(54) Title: NUCLEIC ACID LIGANDS

(57) Abstract

A new class of nucleic acid compounds, referred to as nucleic acid antibodies, have been shown to exist that have a specific binding affinity for three dimensional molecular targets. In a preferred embodiment the ligand components of the nucleic acid antibodies are identified by the method of the invention referred to as the Systematic Evolution of Ligands by Exponential enrichment (SELEX), wherein a candidate mixture of nucleic acids are iteratively enriched in high affinity nucleic acids and amplified for further partitioning.

1
NUCLEIC ACID LIGANDS

This application is a Continuation-in-Part of United States Patent Application Serial No. 07/536,428, filed June 11, 1990, entitled Systematic Evolution of Ligands By Exponential Enrichment.

This work was supported by grants from the United States Government funded through the National Institutes of Health. The U.S. Government has certain rights in this invention.

FIELD OF THE INVENTION

We describe herein a new class of high-affinity nucleic acid ligands that specifically bind a desired target molecule. A method is presented for selecting a nucleic acid ligand that specifically binds any desired target molecule. The method is termed SELEX, an acronym for Systematic Evolution of Ligands by Exponential enrichment. The method of the invention (SELEX) is useful to isolate a nucleic acid ligand for a desired target molecule. The nucleic acid products of the invention are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, nucleic acid products of the invention can have catalytic activity. Target molecules include natural and synthetic polymers, including proteins, polysaccharides, glycoproteins, hormones, receptors and cell surfaces, and small molecules such as drugs, metabolites, cofactors, transition state analogs and toxins.

BACKGROUND OF THE INVENTION

Most proteins or small molecules are not known to specifically bind to nucleic acids. The known protein exceptions are those regulatory proteins such

2

as repressors, polymerases, activators and the like
which function in a living cell to bring about the
transfer of genetic information encoded in the nucleic
acids into cellular structures and the replication of
5   the genetic material. Furthermore, small molecules
such as GTP bind to some intron RNAs.

Living matter has evolved to limit the function
of nucleic acids to a largely informational role. The
Central Dogma, as postulated by Crick, both originally
10   and in expanded form, proposes that nucleic acids
(either RNA or DNA) can serve as templates for the
synthesis of other nucleic acids through replicative
processes that "read" the information in a template
nucleic acid and thus yield complementary nucleic
15   acids. All of the experimental paradigms for genetics
and gene expression depend on these properties of
nucleic acids: in essence, double-stranded nucleic
acids are informationally redundant because of the
chemical concept of base pairs and because replicative
20   processes are able to use that base pairing in a
relatively error-free manner.

The individual components of proteins, the
twenty natural amino acids, possess sufficient chemical
differences and activities to provide an enormous
25   breadth of activities for both binding and catalysis.
Nucleic acids, however, are thought to have narrower
chemical possibilities than proteins, but to have an
informational role that allows genetic information to
be passed from virus to virus, cell to cell, and
30   organism to organism. In this context nucleic acid
components, the nucleotides, must possess only pairs of
surfaces that allow informational redundancy within a
Watson-Crick base pair. Nucleic acid components need
not possess chemical differences and activities
35   sufficient for either a wide range of binding or
catalysis.

However, some nucleic acids found in nature do

participate in binding to certain target molecules and
even a few instances of catalysis have been reported.
The range of activities of this kind is narrow compared
to proteins and more specifically antibodies. For
5   example, where nucleic acids are known to bind to some
protein targets with high affinity and specificity, the
binding depends on the exact sequences of nucleotides
that comprise the DNA or RNA ligand. Thus, short
double-stranded DNA sequences are known to bind to
10   target proteins that repress or activate transcription
in both prokaryotes and eukaryotes. Other short
double-stranded DNA sequences are known to bind to
restriction endonucleases, protein targets that can be
selected with high affinity and specificity. Other
15   short DNA sequences serve as centromeres and telomeres
on chromosomes, presumably by creating ligands for the
binding of specific proteins that participate in
chromosome mechanics. Thus, double-stranded DNA has a
well-known capacity to bind within the nooks and
20   crannies of target proteins whose functions are
directed to DNA binding. Single-stranded DNA can also
bind to some proteins with high affinity and
specificity, although the number of examples is rather
smaller. From the known examples of double-stranded
25   DNA binding proteins, it has become possible to
describe the binding interactions as involving various
protein motifs projecting amino acid side chains into
the major groove of B form double-stranded DNA,
providing the sequence inspection that allows
30   specificity.

       Double-stranded RNA occasionally serves as a
ligand for certain proteins, for example, the
endonuclease RNase III from E. coli. There are more
known instances of target proteins that bind to single-
35   stranded RNA ligands, although in these cases the
single-stranded RNA often forms a complex three-
dimensional shape that includes local regions of

4

intramolecular double-strandedness. The amino-acyl .
tRNA synthetases bind tightly to tRNA molecules with
high specificity. A short region within the genomes of
RNA viruses binds tightly and with high specificity to
the viral coat proteins. A short sequence of RNA binds
to the bacteriophage T4-encoded DNA polymerase, again
with high affinity and specificity. Thus, it is
possible to find RNA and DNA ligands, either double- or
single-stranded, serving as binding partners for
specific protein targets. Most known DNA binding
proteins bind specifically to double-stranded DNA,
while most RNA binding proteins recognize single-
stranded RNA. This statistical bias in the literature
no doubt reflects the present biosphere's statistical
predisposition to use DNA as a double-stranded genome
and RNA as a single-stranded entity in the many roles
RNA plays beyond serving as a genome. Chemically there
is no strong reason to dismiss single-stranded DNA as a
fully able partner for specific protein interactions.

     RNA and DNA have also been found to bind to
smaller target molecules. Double-stranded DNA binds to
various antibiotics, such as actinomycin D. A specific
single-stranded RNA binds to the antibiotic
thiostreptone; specific RNA sequences and structures
probably bind to certain other antibiotics, especially
those whose functions is to inactivate ribosomes in a
target organism. A family of evolutionarily related
RNAs binds with specificity and decent affinity to
nucleotides and nucleosides (Bass, B. and Cech, T.
(1984) Nature 308:820-826) as well as to one of the
twenty amino acids (Yarus, M. (1988) Science 240:1751-
1758). Catalytic RNAs are now known as well, although
these molecules perform over a narrow range of chemical
possibilities, which are thus far related largely to
phosphodiester transfer reactions and hydrolysis of
nucleic acids.

     Despite these known instances, the great

5

majority of proteins and other cellular components are
thought not to bind to nucleic acids under
physiological conditions and such binding as may be
observed is non-specific. Either the capacity of

5   nucleic acids to bind other compounds is limited to the
relatively few instances enumerated supra, or the
chemical repertoire of the nucleic acids for specific
binding is avoided (selected against) in the structures
that occur naturally. The present invention is

10   premised on the inventors' fundamental insight that
nucleic acids as chemical compounds can form a
virtually limitless array of shapes, sizes and
configurations, and are capable of a far broader
repertoire of binding and catalytic functions than

15   those displayed in biological systems.

The chemical interactions have been explored in
cases of certain known instances of protein-nucleic
acid binding. For example, the size and sequence of
the RNA site of bacteriophage R17 coat protein binding

20   has been identified by Uhlenbeck and coworkers. The
minimal natural RNA binding site (21 bases long) for
the R17 coat protein was determined by subjecting
variable-sized labeled fragments of the mRNA to
nitrocellulose filter binding assays in which protein-

25   RNA fragment complexes remain bound to the filter
(Carey et al. (1983) Biochemistry 22:2601). A number
of sequence variants of the minimal R17 coat protein
binding site were created in vitro in order to
determine the contributions of individual nucleic acids

30   to protein binding (Uhlenbeck et al. (1983) J. Biomol.
Structure Dynamics 1:539 and Romaniuk et al. (1987)
Biochemistry 26:1563). It was found that the
maintenance of the hairpin loop structure of the
binding site was essential for protein binding but, in

35   addition, that nucleotide substitutions at most of the
single-stranded residues in the binding site, including
a bulged nucleotide in the hairpin stem, significantly

6

affected binding. In similar studies, the binding of
bacteriophage Qβ coat protein to its translational
operator was examined (Witherell and Uhlenbeck (1989)
Biochemistry 28:71). The Qβ coat protein RNA binding
5   site was found to be similar to that of R17 in size,
and in predicted secondary structure, in that it
comprised about 20 bases with an 8 base pair hairpin
structure which included a bulged nucleotide and a 3
base loop. In contrast to the R17 coat protein binding
10   site, only one of the single-stranded residues of the
loop is essential for binding and the presence of the
bulged nucleotide is not required. The protein-RNA
binding interactions involved in translational
regulation display significant specifity.

15   Nucleic acids are known to form secondary and
tertiary structures in solution. The double-stranded
forms of DNA include the so-called B double-helical
form, Z-DNA and superhelical twists (Rich, A. et al.
(1984) Ann. Rev. Biochem. 53:791-846). Single-stranded
20   RNA forms localized regions of secondary structure such
as hairpin loops and pseudoknot structures (Schimmel,
P. (1989) Cell 58:9-12). However, little is known
concerning the effects of unpaired loop nucleotides on
stability of loop structure, kinetics of formation and
25   denaturation, thermodynamics, and almost nothing is
known of tertiary structures and three dimensional
shape, nor of the kinetics and thermodynamics of
tertiary folding in nucleic acids (Tuerk, C. et al.
(1988) Proc. Natl. Acad. Sci. USA 85:1364-1368).

30   A type of in vitro evolution was reported in
replication of the RNA bacteriophage Qβ. Mills, D.R.
et al. (1967) Proc. Natl. Acad. Sci USA 58:217-224;
Levinsohn, R. and Spiegleman, S. (1968) Proc. Natl.
Acad. Sci. USA 60:866-872; Levisohn, R. and Spiegelman
35   S. (1969) Proc. Natl. Acad. Sci. USA 63:805-811;
Saffhill, R. et al. (1970) J. Mol. Biol. 51:531-539;
Kacian, D.L. et al. (1972) Proc. Natl. Acad. Sci. USA

<u>69</u>:3038-3042; Mills, D.R. et al. (1973) Science
<u>180</u>:916-927. The phage RNA serves as a poly-cistronic
messenger RNA directing translation of phage-specific
proteins and also as a template for its own replication
catalyzed by Qβ RNA replicase. This RNA replicase was
shown to be highly specific for its own RNA templates.
During the course of cycles of replication <u>in vitro</u>
small variant RNAs were isolated which were also
replicated by Qβ replicase. Minor alterations in the
conditions under which cycles of replication were
performed were found to result in the accumulation of
different RNAs, presumably because their replication
was favored under the altered conditions. In these
experiments, the selected RNA had to be bound
efficiently by the replicase to initiate replication
and had to serve as a kinetically favored template
during elongation of RNA. Kramer <u>et al</u>. (1974) J. Mol.
Biol. <u>89</u>:719 reported the isolation of a mutant RNA
template of Qβ replicase, the replication of which was
more resistant to inhibition by ethidium bromide than
the natural template. It was suggested that this
mutant was not present in the initial RNA population
but was generated by sequential mutation during cycles
of <u>in vitro</u> replication with Qβ replicase. The only
source of variation during selection was the intrinsic
error rate during elongation by Qβ replicase. In these
studies what was termed "selection" occurred by
preferential amplification of one or more of a limited
number of spontaneous variants of an initially
homogenous RNA sequence. There was no selection of a
desired result, only that which was intrinsic to the
mode of action of Qβ replicase.

Joyce and Robertson (Joyce (1989) in <u>RNA:
Catalysis, Splicing, Evolution</u>, Belfort and Shub
(eds.), Elsevier, Amsterdam pp. 83-87; and Robertson
and Joyce (1990) Nature <u>344</u>:467) reported a method for
identifying RNAs which specifically cleave single-

8

stranded DNA.   The selection for catalytic activity was
based on the ability of the ribozyme to catalyze the
cleavage of a substrate ssRNA or DNA at a specific
position and transfer the 3'-end of the substrate to
the 3'-end of the ribozyme.   The product of the desired
reaction was selected by using an oligodeoxynucleotide
primer which could bind only to the completed product
across the junction formed by the catalytic reaction
and allowed selective reverse transcription of the
ribozyme sequence.   The selected catalytic sequences
were amplified by attachment of the promoter of T7 RNA
polymerase to the 3'-end of the cDNA, followed by
transcription to RNA.   The method was employed to
identify from a small number of ribozyme variants the
variant that was most reactive for cleavage of a
selected substrate.   Only a limited array of variants
was testable, since variation depended upon single
nucleotide changes occurring during amplification.

          The prior art has not taught or suggested more
than a limited range of chemical functions for nucleic
acids in their interactions with other substances: as
targets for protein ligands evolved to bind certain
specific olignocleotide sequences; more recently, as
catalysts with a limited range of activities.   Prior
"selection" experiments have been limited to a narrow
range of variants of a previously described function.
Now, for the first time, it will be understood that the
nucleic acids are capable of a vastly broad range of
functions and the methodology for realizing that
capability is disclosed herein.

SUMMARY OF THE INVENTION

          The present invention provides a class of
products which are nucleic acid molecules, each having
a unique sequence, each of which has the property of
binding specifically to a desired target compound or
molecule.   Each compound of the invention is a specific

SUBSTITUTE SHEET

9

ligand of a given target molecule. The invention is based on the unique insight that nucleic acids have sufficient capacity for forming a variety of two- and three-dimensional structures and sufficient chemical versatility available within their monomers to act as ligands (form specific binding pairs) with virtually any chemical compound, whether monomeric or polymeric. Molecules of any size can serve as targets. Most commonly, and preferably, for therapeutic applications, binding takes place in aqueous solution at conditions of salt, temperature and pH near acceptable physiological limits.

The invention also provides a method which is generally applicable to make a nucleic acid ligand for any desired target. The method involves selection from a mixture of candidates and step-wise iterations of structural improvement, using the same general selection theme, to achieve virtually any desired criterion of binding affinity and selectivity. Starting from a mixture of nucleic acids, preferably comprising a segment of randomized sequence, the method, termed SELEX herein, includes steps of contacting the mixture with the target under conditions favorable for binding, partitioning unbound nucleic acids from those nucleic acids which have bound to target molecules, dissociating the nucleic acid-target pairs, amplifying the nucleic acids dissociated from the nucleic acid-target pairs to yield a ligand-enriched mixture of nucleic acids, then reiterating the steps of binding, partitioning, dissociating and amplifying through as many cycles as desired.

While not bound by a theory of preparation, SELEX is based on the inventors' insight that within a nucleic acid mixture containing a large number of possible sequences and structures there is a wide range of binding affinities for a given target. A nucleic

acid mixture comprising, for example a 20 nucleotide
randomized segment can have 4^{20} candidate
possibilities. Those which have the higher affinity
constants for the target are most likely to bind.
After partitioning, dissociation and amplification, a
second nucleic acid mixture is generated, enriched for
the higher binding affinity candidates. Additional
rounds of selection progressively favor the best
ligands until the resulting nucleic acid mixture is
predominantly composed of only one or a few sequences.
These can then be cloned, sequenced and individually
tested for binding affinity as pure ligands.

Cycles of selection and amplification are
repeated until a desired goal is achieved. In the most
general case, selection/amplification is continued
until no significant improvement in binding strength is
achieved on repetition of the cycle. The iterative
selection/ amplification method is sensitive enough to
allow isolation of a single sequence variant in a
mixture containing at least 65,000 sequence variants.
The method is even capable of isolating a small number
of high affinity sequences in a mixture containing 10^{14}
sequences. The method could, in principle, be used to
sample as many as about 10^{18} different nucleic acid
species. The nucleic acids of the test mixture
preferably include a randomized sequence portion as
well as conserved sequences necessary for efficient
amplification. Nucleic acid sequence variants can be
produced in a number of ways including synthesis of
randomized nucleic acid sequences and size selection
from randomly cleaved cellular nucleic acids. The
variable sequence portion may contain fully or
partially random sequence; it may also contain
subportions of conserved sequence incorporated with
randomized sequence. Sequence variation in test
nucleic acids can be introduced or increased by
mutagenesis before or during the

selection/amplification iterations.

In one embodiment of the present invention, the selection process is so efficient at isolating those nucleic acid ligands that bind most strongly to the
5   selected target, that only one cycle of selection and amplification is required. Such an efficient selection may occur, for example, in a chromatographic-type process wherein the ability of nucleic acids to associate with targets bound on a column operates in
10   such a manner that the column is sufficiently able to allow separation and isolation of the highest affinity nucleic acid ligands.

In many cases, it is not necessarily desirable to perform the iterative steps of SELEX until a single
15   nucleic acid ligand is identified. The target-specific nucleic acid ligand solution may include a family of nucleic acid structures or motifs that have a number of conserved sequences and a number of sequences which can be substituted or added without significantly effecting
20   the affinity of the nucleic acid ligands to the target. By terminating the SELEX process prior to completion, it is possible to determine the sequence of a number of members of the nucleic acid ligand solution family, which will allow the determination of a comprehensive
25   description of the nucleic acid ligand solution.

After a description of the nucleic acid ligand family has been resolved by SELEX, in certain cases it may be desirable to perform a further series of SELEX that is tailored by the information received during the
30   SELEX experiment. In one embodiment, the second series of SELEX will fix those conserved regions of the nucleic acid ligand family while randomizing all other positions in the ligand structure. In an alternate embodiment, the sequence of the most representative
35   member of the nucleic acid ligand family may be used as the basis of a SELEX process wherein the original pool of nucleic acid sequences is not completely randomized

but contains biases towards the best known ligand. By
these methods it is possible to optimize the SELEX
process to arrive at the most preferred nucleic acid
ligands.

A variety of nucleic acid primary, secondary
and tertiary structures are known to exist. The
structures or motifs that have been shown most commonly
to be involved in non-Watson-Crick type interactions
are referred to as hairpin loops, symmetric and
asymmetric bulges, psuedoknots and myriad combinations
of the same. Almost all known cases of such motifs
suggest that they can be formed in a nucleic acid
sequence of no more than 30 nucleotides. For this
reason, it is preferred that SELEX procedures with
contiguous randomized segments be initiated with
nucleic acid sequences containing a randomized segment
of between about 20-50 nucleotides, and in the most
preferred embodiments between 25 and 40 nucleotides.
This invention includes solutions comprising a mixture
of between about 10^9 to 10^{18} nucleic acid sequences
having a contiguous randomized sequence of at least
about 15 nucleotides in length. In the preferred
embodiment, the randomized section of sequences is
flanked by fixed sequences that facilitate the
amplification of the ligands.

In the case of a polymeric target, such as a
protein, the ligand affinity can be increased by
applying SELEX to a mixture of candidates comprising a
first selected sequence and second randomized sequence.
The sequence of the first selected ligand associated
with binding or subportions thereof can be introduced
into the randomized portion of the nucleic acids of a
second test mixture. The SELEX procedure is repeated
with the second test mixture to isolate a second
nucleic acid ligand, having two sequences selected for
binding to the target, which has increased binding
strength or increased specificity of binding compared

13

to the first nucleic acid ligand isolated. The
sequence of the second nucleic acid ligand associated
with binding to the target can then be introduced into
the variable portion of the nucleic acids of a third
test mixture which, after cycles of SELEX results in a
third nucleic acid ligand. These procedures can be
repeated until a nucleic acid ligand of a desired
binding strength or a desired specificity of binding to
the target molecule is achieved. The process of
iterative selection and combination of nucleic acid
sequence elements that bind to a selected target
molecule is herein designated "walking," a term which
implies the optimized binding to other accessible areas
of a macromolecular target surface or cleft, starting
from a first binding domain. Increasing the area of
binding contact between ligand and target can increase
the affinity constant of the binding reaction. These
walking procedures are particularly useful for the
isolation of nucleic acid antibodies which are highly
specific for binding to a particular target molecule.

A variant of the walking procedure employs a
non-nucleic acid ligand termed "anchor" which binds to
the target molecule as a first binding domain. (See
Fig. 9.) This anchor molecule can in principle be any
non-nucleic acid molecule that binds to the target
molecule and which can be covalently linked directly or
indirectly to a nucleic acid. When the target molecule
is an enzyme, for example, the anchor molecule can be
an inhibitor or substrate of that enzyme. The anchor
can also be an antibody or antibody fragment specific
for the target. The anchor molecule is covalently
linked to a nucleic acid oligomer of known sequence to
produce a bridging molecule. The oligomer is
preferably comprised of a minimum of about 3 - 10
bases. A test mixture of candidate nucleic acids is
then prepared which includes a randomized portion and a
sequence complementary to the known sequence of the

14

bridging molecule. The bridging molecule is complexed
to the target molecule. SELEX is then applied to
select nucleic acids which bind to the complex of the
bridging molecule and the target molecule. Nucleic
acid ligands which bind to the complex are isolated.
Walking procedures as described above can then be
applied to obtain nucleic acid ligands with increased
binding strength or increased specificity of binding to
the complex. Walking procedures could employ
selections for binding to the complex or the target
itself. This method is particularly useful to isolate
nucleic acid ligands which bind at a particular site
within the target molecule. The complementary sequence
in the test mixture acts to ensure the isolation of
nucleic acid sequences which bind to the target
molecule at or near the binding site of the bridging
molecule. If the bridging molecule is derived from an
inhibitor of the target molecule, this method is likely
to result in a nucleic acid ligand which inhibits the
function of the target molecule. It is particularly
useful, for example, for the isolation of nucleic acids
which will activate or inhibit protein function. The
combination of ligand and target can have a new or
enhanced function.

        The nucleic acid ligands of the present
invention may contain a plurality of ligand components.
As described above, nucleic acid ligands derived by
walking procedures may be considered as having more
than one nucleic acid ligand component. This invention
also includes nucleic acid antibodies that are
constructed based on the results obtained by SELEX
while not being identical to a nucleic acid ligand
identified by SELEX. For example, a nucleic acid
antibody may be constructed wherein a plurality of
identical ligand structures are made part of a single
nucleic acid. In another embodiment, SELEX may
identify more than one family of nucleic acid ligands

SUBSTITUTE SHEET

15

to a given target. In such case, a single nucleic acid
antibody may be constructed containing a plurality of
different ligand structures. SELEX experiments also
may be performed wherein fixed identical or different
ligand structures are joined by random nucleotide
regions and/or regions of varying distance between the
fixed ligand structures to identify the best nucleic
acid antibodies.

Screens, selections or assays to assess the
effect of binding of a nucleic acid ligand on the
function of the target molecule can be readily combined
with the SELEX methods. Specifically, screens for
inhibition or activation of enzyme activity can be
combined with the SELEX methods.

In more specific embodiments, the SELEX method
provides a rapid means for isolating and identifying
nucleic acid ligands which bind to proteins, including
both nucleic acid-binding proteins and proteins not
known to bind nucleic acids as part of their biological
function. Nucleic acid-binding proteins include among
many others polymerases and reverse transcriptases.
The methods can also be readily applied to proteins
which bind nucleotides, nucleosides, nucleotide co-
factors and structurally related molecules.

In another aspect, the present invention
provides a method for detecting the presence or absence
of, and/or measuring the amount of a target molecule in
a sample, which method employs a nucleic acid ligand
which can be isolated by the methods described herein.
Detection of the target molecule is mediated by its
binding to a nucleic acid ligand specific for that
target molecule. The nucleic acid ligand can be
labeled, for example radiolabled, to allow qualitative
or quantitative detection. The detection method is
particularly useful for target molecules which are
proteins. The method is more particularly useful for
detection of proteins which are not known to bind

nucleic acids as part of their biological function. Thus, nucleic acid ligands of the present invention can be employed in diagnostics in a manner similar to conventional antibody-based diagnostics. One advantage of nucleic acid ligands over conventional antibodies in such detection method and diagnostics is that nucleic acids are capable of being readily amplified _in_ _vitro_, for example, by use of PCR amplification or related methods. Another advantage is that the entire SELEX process is carried out _in_ _vitro_ and does not require immunizing test animals. Furthermore, the binding affinity of nucleic acid ligands can be tailored to the user's needs.

Nucleic acid ligands of small molecule targets are useful as diagnostic assay reagents and have therapeutic uses as sequestering agents, drug delivery vehicles and modifiers of hormone action. Catalytic nucleic acids are selectable products of this invention. For example, by selecting for binding to transition state analogs of an enzyme catalyzed reaction, catalytic nucleic acids can be selected.

In yet another aspect, the present invention provides a method for modifying the function of a target molecule using nucleic acid ligands which can be isolated by SELEX. Nucleic acid ligands which bind to a target molecule are screened to select those which specifically modify function of the target molecule, for example to select inhibitors or activators of the function of the target molecule. An amount of the selected nucleic acid ligand which is effective for modifying the function of the target is combined with the target molecule to achieve the desired functional modification. This method is particularly applicable to target molecules which are proteins. A particularly useful application of this method is to inhibit protein function, for example to inhibit receptor binding to an effector or to inhibit enzyme catalysis. In this case,

an amount of the selected nucleic acid molecule which
is effective for target protein inhibition is combined
with the target protein to achieve the desired
inhibition.

5

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a diagram of the ribonucleotide
sequence of a portion of the gene 43 messenger RNA
which encodes the bacteriophage T4 DNA polymerase.
10   Shown is the sequence in the region known to bind to
gp43. The bold-faced capitalized letters indicate the
extent of the information required for binding of gp43.
The eight base-pair loop was replaced by randomized
sequence to yield a candidate population for SELEX.
15   Figure 2 is a schematic diagram of the SELEX
process as exemplified for selecting loop sequence
variants for RNAs that bind to T4 DNA polymerase
(gp43). A DNA template for preparation of a test
mixture of RNAs was prepared as indicated in step a by
20   ligation of oligomers 3, 4 and 5, whose sequences are
given in Table 1 infra. Proper ligation in step a was
assured by hybridization with oligomers 1 and 2, which
have complementary sequence (given in Table 1) that
bridges oligomers 3 and 4 and 4 and 5, respectively.
25   The resultant 110-base long template was gel-purified,
annealed to oligo 1 and was used in vitro transcription
reactions (Miligan et al. (1987) Nucl. Acids Res.
15:8783-8798) to produce an initial RNA mixture
containing randomized sequences of the 8-base loop,
30   step b. The resultant transcripts were gel-purified
and subjected to selection on nitrocellulose filters
for binding to gp43 (step c), as described in Example
1. Selected RNAs were amplified in a three step
process: (d) cDNA copies of the selected RNAs were
35   made by reverse transcriptase synthesis using oligo 5
(Table 1) as a primer; (e) cDNAs were amplified using
Taq DNA polymerase chain extension of oligo 1 (Table

18

1), which carries essential T7 promoter sequences, and oligo 5 (Table 1) as described in Innis et al. (1988) Proc. Natl. Acad. Sci. USA 85:9436; and (f) double-stranded DNA products of amplification were transcribed
5   in vitro. The resultant selected amplified RNAs were used in the next round of selection.

Figure 3 is a composite of autoradiographs of electrophoresed batch sequencing reactions of the in vitro transcripts derived from SELEX for binding of RNA
10 loop variants to gp43. The figure indicates the change in loop sequence components as a function of number of selection cycles (for 2, 3 and 4 cycles) for selection conditions of experiment B in which the concentration of gp43 was 3×10^{-8} M and the concentration of RNA was
15 about 3×10^{-5} M in all selection cycles. Sequencing was performed as described in Gauss et al. (1987) Mol. Gen. Genet. 206:24-34.

Figure 4 is a composite of autoradiographs of batch RNA sequences of those RNAs selected from the
20 fourth round of SELEX amplification for binding of RNA loop variants to gp43 employing different binding conditions. In experiment A gp43 concentration was 3×10^{-8} M and RNA concentration was about 3×10^{-7} M. In experiment B, gp43 was 3×10^{-8} M and RNA was about $3 \times$
25 10^{-5} M. In experiment C, gp43 was 3×10^{-7} M and RNA was about 3×10^{-5} M.

Figure 5 is a composite of autoradiographs of three sequencing gels for loop variants selected for binding to gp43 under the selection conditions of
30 experiment B (see Example 1). The left hand sequence gel is the batch sequencing of selected RNAs after the fourth round of selection/amplification. The middle and right hand sequence gels are double-stranded DNA sequencing gels of two clonal isolates derived form the
35 batch RNAs. The batch of RNA selected is composed of two major variants, one of which was the wild-type sequence (middle sequence gel), and a novel sequence

SUBSTITUTE SHEET

(right hand gel).

Figure 6 is a graph of percent RNA bound to
gp43 as a function of gp43 concentration for different
selected RNA loop sequence variants and for RNA with a
randomized loop sequence. Binding of the wild-type
loop sequence AAUAACUC is indicated as open circles,
solid line; major variant loop sequence AGCAACCU as
"x," dotted line; minor variant loop sequence AAUAACUU
as open squares, solid line; minor variant loop
sequence AAUGACUC as solid circles, dotted line; minor
variant loop sequence AGCGACCU as crosses, dotted line;
and binding of the randomized mixture (NNNNNNNN) of
loop sequences as open circles, dotted line.

Figure 7 is a pictorial summary of results
achieved after four rounds of SELEX to select a novel
gp43 binding RNA from a candidate population randomized
in the eight base-pair loop. SELEX did not yield the
"apparent" consensus expected from the batch sequences
shown in Figure 4, but instead yielded wild type and a
single major variant in about equal proportions and
three single mutants. The frequencies of each species
out of twenty cloned isolates are shown together with
the approximate affinity constants (Kd) for each, as
derived from filter binding assays shown in Figure 6.

Figure 8 is a series of diagrams showing
synthesis of candidate nucleic acid ligands using the
enzymes terminal transferase (TDT) and DNA polymerase
(DNA pol). A 5' primer or primary ligand sequence is
provided with a tail of randomized sequence by
incubating with terminal transferase in the presence of
the four deoxynucleotide triphosphates (dNTPs).
Homopolymer tailing of the randomized segment, using
the same enzyme in the presence of a single
deoxynucleotide triphosphate (e.g. dCTP) provides an
annealing site for poly-G tailed 3' primer. After
annealing, the double-stranded molecule is completed by
the action of DNA polymerase. The mixture can be

further amplified, if desired, by the polymerase chain reaction.

Fig. 9 is a diagram showing a process using SELEX to select a large nucleic acid ligand having two spatially separate binding interactions with a target protein. The process is termed "walking" since it includes two stages, the second being an extension of the first. The upper part of the figure depicts a target ("protein of interest") with a bound nucleic acid ligand selected by a first round of SELEX ("evolved primary ligand") bound to the protein at a first binding site. A reaction catalyzed by terminal transferase extends the length of the evolved primary ligand and generates a new set of randomized sequence candidates having a conserved region containing the primary ligand. The lower part of the figure depicts the result of a second round of SELEX based upon improved binding that results from the secondary ligand interaction at the secondary binding site of the protein. The terms "primary" and "secondary" are merely operative terms that do not imply that one has higher affinity than the other.

Figures 10 and 11 are diagrams of a selection process using SELEX in two stages. In Figure 10, SELEX is applied to select ligands that bind to secondary binding sites on a target complexed with a bridging oligonucleotide connected to a specific binder, e.g., inhibitor of the target protein. The bridging oligonucleotide acts as a guide to favor selection of ligands that bind to accessible secondary binding sites. In Figure 11, a second SELEX is applied to evolve ligands that bind at both the secondary sites originally selected for and the primary target domain. The nucleic acids thereby evolved will bind very tightly, and may themselves act as inhibitors of the target protein or to compete against inhibitors or substrates of the target protein.

SUBSTITUTE SHEET

21

Figure 12 shows the sequence and placement of oligomers used to construct the candidate mixture used in Example 2. The top line shows the sequences of oligomers 1b and 2b from left to right, respectively (see Table 2 _infra_). The second line shown, from left to right, the sequences of oligomers 3b, 4b and 5b (Table 2). Proper ligation of the oligomers was assured by hybridization with oligomers 1b and 2b, whose sequences are complementary. The resultant ligated template was gel-purified, annealed to oligomer 1b and used in an _in vitro_ transcription reaction (Milligan et al. (1987)) to produce an RNA candidate mixture, shown in the last line of the figure, labeled "_in vitro_ transcript." The candidate mixture contained a 32 nucleotide randomized segment, as shown.

Figure 13 shows a hypothetical RNA sequence containing a variety of secondary structures that RNA are known to undertake. Included are: A hairpin loops, B bulges, C asymmetric bulges, and D pseudoknots.

Figure 14 shows nitrocellulose filter binding assays of ligand affinity for HIV-RT. Shown is the percent of input RNA that is bound to the nitrocellulose filter with varying concentrations of HIV-RT.

Figure 15 shows additional nitrocellulose filter binding assays of ligand affinity for HIV-RT.

Figure 16 shows information boundary determination for HIV-1 RT ligands 1.1 and 1.3a a) 3' boundary determination. RNAs were 5' end labeled, subjected to partial alkaline hydrolysis and selection on nitrocellulose filters, separated on a denaturing 8% polyacrylamide gel and autoradiographed. Approximately 90 picomoles of labeled RNA and 80 picomoles of HIV-1 RT were mixed in 0.5, 2.5, and 5 mls of buffer and incubated for 5 minutes at 37° C prior to washing through a nitrocellulose filter. The eluated RNAs are

SUBSTITUTE SHEET

shown under the final concentrations of HIV-1 RT used
in each experiment. Also shown are the products of a
partial RNase T1 digest which allows identification of
the information boundary on the adjacent sequence as

5   shown by arrows b) 5' boundary determination. The 5'
boundary was determined in a) under the same conditions
listed above.

Figure 17 shows the inhibition of HIV-1 RT by
RNA ligand 1.1. A series of three-fold dilutions of

10   32N candidate mixture RNA and ligand 1.1 RNA ranging in
final reaction concentration for 10 micromolar to 4.6
nanomolar and pre-mixed with HIV-RT and incubated for 5
minutes at 37°C in 6μL of 200 mM KOAc, 50 mM Tris-HCl,
pH 7.7, 10 mM dithiothreitol, 6 mM Mg (OAc)$_2$, and 0.4

15   mM NTPS. In a separate tube RNA template (transcribed
from a PCR product of a T7-1 obtained from U.S.
Biochemical Corp. using oligos 7 and 9) and labeled
oligo 9 were mixed and heated at 95°C for one minute
and cooled on ice for 15 minutes in 10 mM Tris-HCl, pH

20   7, 0.1 mM EDTA. Four μl of this template was added to
each 6μl enzyme-inhibitor mixture to start the reaction
which was incubated for a further 5 minutes at 37°C and
then stopped. The final concentration of HIV-1 RT was
16 nanomolar, of RNA template was 13 nanomolar, and of

25   labeled primer was 150 nanomolar in all reactions. The
extension products of each reaction are shown.

Figure 18 shows a comparisons of HIV-1 RT
inhibition by ligand 1.1 to effects on MMLV RT and AMV
RT. Experiments were performed as in Figure 17 except

30   that 5-fold dilutions of inhibitor were prepared with
the resultant concentrations as shown. The
concentrations of each RT were normalized to that of
HIV-RT by dilutions and comparison of gel band
intensity with both Coomassie blue and silver stains,

35   Biorad protein concentration assays, and activity
assays.

Figure 19 shows the consensus sequences of

23

selected hairpins representing the R-17 coat protein
ligand solution. The nucleotide representation at each
position is indicated in grids. The column headed
"bulge" represents the number of clones with an
5   extra-helical nucleotide on one or both sides of the
stem between the corresponding stem base-pairs. The
column headed "end" represents the number of clones
whose hairpin terminated at the previous base-pair.

Figure 20 shows a binding curve of 30N bulk RNA
10   for bradykinin. Anaylsis was done using spin columns;
10 mM KOAc, 10 mM DEM, pH 7.5; RNA concentration 1.5 x
10^{-8}M.

Figure 21 shows templates for use in the
generation of candidate mixtures that are enriched in
15   certain structural motifs. Template A is designed to
enrich the candidate mixture in hairpin loops.
Template B is designed to enrich the candidate mixture
in pseudoknots.

Figure 22 is a schematic diagram of stem-loop
20   arrangements for Motifs I and II of the HIV-rev ligand
solution. The dotted lines in stems 1 and 2 between
loops 1 and 3 indicate potential base-pairs.

Figure 23 shows the folded secondary structures
of rev ligand subdomains of isolates 6a, 1a, and 8 to
25   show motifs I, II and III respectively. Also shown for
comparison is the predicted fold of the wild type RRE
RNA.

Figure 24 is a graph of percent of input counts
bound to a nitrocellulose filter with various
30   concentrations of HIV rev protein. Also shown are the
binding curves of the 32N starting population (#) and
of the evolved population after 10 rounds (P) and of
the wild type RRE sequence transcribed from a template
composed of oligos 8 and 9(W).
35   Figure 25 is a comparison of Motif I(a) rev
ligands. Paramenters are as in Figure 24. Also
included is the binding curve of the "consensus"

24

construct (C).

Figure 26 is a comparison of Motif I(b) rev ligands. Parameters are as in Figure 24.

Figure 27 is a comparison of Motif II rev ligands. Parameters are as in Figure 24.

Figure 28 is a comparison of Motif III rev ligands. Parameters as in Figure 24.

Figure 29 shows the consensus nucleic acid ligand solution to HIV rev referred to as Motif I.

Figure 30 shows the consensus nucleic acid ligand solution to HIV rev referred to as Motif II.

Figure 31 is a schematic representation of a pseudoknot. The pseudoknot consists of two stems and three loops, referred to herein as stems S_1 and S_2 and loops 1, 2 and 3.

DETAILED DESCRIPTION OF THE INVENTION

The following terms are used herein according to the definitions.

Nucleic acid means either DNA, RNA, single-stranded or double-stranded and any chemical modifications thereof, provided only that the modification does not interfere with amplification of selected nucleic acids. Such modifications include, but are not limited to, modifications at cytosine exocyclic amines, substitution of 5-bromo-uracil, backbone modifications, methylations, unusual base-pairing combinations and the like.

Ligand means a nucleic acid that binds another molecule (target). In a population of candidate nucleic acids, a ligand is one which binds with greater affinity than that of the bulk population. In a candidate mixture there can exist more than one ligand for a given target. The ligands can differ from one another in their binding affinities for the target molecule.

Candidate mixture is a mixture of nucleic acids

SUBSTITUTE SHEET

of differing sequence, from which to select a desired ligand. The source of a candidate mixture can be from naturally-occurring nucleic acids or fragments thereof, chemically synthesized nucleic acids, enzymically synthesized nucleic acids or nucleic acids made by a combination of the foregoing techniques.

Target molecule means any compound of interest for which a ligand is desired. A target molecule can be a protein, peptide, carbohydrate, polysaccharide, glycoprotein, hormone, receptor, antigen, antibody, virus, substrate, metabolite, transition state analog, cofactor, inhibitor, drug, dye, nutrient, growth factor, etc., without limitation.

Partitioning means any process whereby ligands bound to target molecules, termed ligand-target pairs herein, can be separated from nucleic acids not bound to target molecules. Partitioning can be accomplished by various methods known in the art. Nucleic acid-protein pairs can be bound to nitrocellulose filters while unbound nucleic acids are not. Columns which specifically retain ligand-target pairs (or specifically retain bound ligand complexed to an attached target) can be used for partitioning. Liquid-liquid partition can also be used as well as filtration gel retardation, and density gradient centrifugation. The choice of partitioning method will depend on properties of the target and of the ligand-target pairs and can be made according to principles and properties known to those of ordinary skill in the art.

Amplifying means any process or combination of process steps that increases the amount or number of copies of a molecule or class of molecules. Amplifying RNA molecules in the disclosed examples was carried out by a sequence of three reactions: making cDNA copies of selected RNAs, using polymerase chain reaction to increase the copy number of each cDNA, and transcribing the cDNA copies to obtain RNA molecules having the same

sequences as the selected RNAs. Any reaction or
combination of reactions known in the art can be used
as appropriate, including direct DNA replication,
direct RNA amplification and the like, as will be

5   recognized by those skilled in the art. The
amplification method should result in the proportions
of the amplified mixture being essentially
representative of the proportions of different
sequences in the initial mixture.

10   Specific binding is a term which is defined on
a case-by-case basis. In the context of a given
interaction between a given ligand and a given target,
a binding interaction of ligand and target of higher
affinity than that measured between the target and the

15   candidate ligand mixture is observed. In order to
compare binding affinities, the conditions of both
binding reactions must be the same, and should be
comparable to the conditions of the intended use. For
the most accurate comparisons, measurements will be

· 20   made that reflect the interaction between ligand as a
whole and target as a whole. The nucleic acid ligands
of the invention can be selected to be as specific as
required, either by establishing selection conditions
that demand the requisite specificity during SELEX, or

25   by tailoring and modifying the ligand through "walking"
and other modifications using interactions of SELEX.
            Randomized is a term used to described a
segment of a nucleic acid having, in principle any
possible sequence over a given length. Randomized

30   sequences will be of various lengths, as desired,
ranging from about eight to more than 100 nucleotides.
The chemical or enzymatic reactions by which random
sequence segments are made may not yield mathematically
random sequences due unknown biases or nucleotide

35   preferences that may exist. The term "randomized" is
used instead of "random" to reflect the possibility of
such deviations from

non-ideality. In the techniques presently known, for
example sequential chemical synthesis, large deviations
are not known to occur. For short segments of 20
nucleotides or less, any minor bias that might exist
would have negligible consequences. The longer the
sequences of a single synthesis, the greater the effect
of any bias.

A bias may be deliberately introduced into
randomized sequence, for example, by altering the molar
ratios of precursor nucleoside (or deoxynucleoside)
triphosphates of the synthesis reaction. A deliberate
bias may be desired, for example, to approximate the
proportions of individual bases in a given organism, or
to affect secondary structure.

SELEXION refers to a mathematical analysis and
computer simulation used to demonstrate the powerful
ability of SELEX to identify nucleic acid ligands and
to predict which variations in the SELEX process have
the greatest impact on the optimization of the process.
SELEXION is an acronym for Systematic Evolution of
Ligands by EXponential enrichment with Integrated
Optimization by Nonlinear analysis.

Nucleic acid antibodies is a term used to refer
to a class of nucleic acid ligands that are comprised
of discrete nucleic acid structures or motifs that
selectively bind to target molecules. Nucleic acid
antibodies may be made up of double or single stranded
RNA or DNA. The nucleic acid antibodies are
synthesized, and in a preferred embodiment are
constructed based on a ligand solution or solutions
received for a given target by the SELEX process. In
many cases, the nucleic acid antibodies of the present
invention are not naturally occurring in nature, while
in other situations they may have significant
similarity to a naturally occurring nucleic acid
sequence.

The nucleic acid antibodies of the present

invention include all nucleic acids having a specific
binding affinity for a target, while not including the
cases when the target is a polynucleotide which binds
to the nucleic acid through a mechanism which

5   predominantly depends on Watson/Crick base pairing or
triple helix agents (See, Riordan, M. et al. (1991)
Nature 350:442-443); provided, however, that when the
nucleic acid antibody is double-stranded DNA, the
target is not a naturally occuring protein whose

10   physiological function depends on specific binding to
double-stranded DNA.

RNA motifs is a term generally used to describe
the secondary or tertiary structure of RNA molecules.
The primary sequence of an RNA is a specific string of

15   nucleotides (A, C, G or U) in one dimension. The
primary sequence does not give information on first
impression as to the three dimensional configuration of
the RNA, although it is the primary sequence that
dictates the three dimensional configuration. In

20   certain cases, the ligand solution obtained after
performing SELEX on a given target may best be
represented as a primary sequence. Although
conformational information pertaining to such a ligand
solution is not always ascertainable based on the

25   results obtained by SELEX, the representation of a
ligand solution as a primary sequence shall not be
interpreted as disclaiming the existence of an integral
tertiary structure.

The secondary structure of an RNA motif is

30   represented by contact in two dimensions between
specific nucleotides. The most easily recognized
secondary structure motifs are comprised of the
Watson/Crick basepairs A:U and C:G. Non-Watson/Crick
basepairs, often of lower stability, have been

35   recognized, and include the pairs G:U, A:C, G:A,and
U:U. (Base pairs are shown once; in RNA molecules the
base pair X:Y by convention represents a sequence in

SUBSTITUTE SHEET

29

which X is 5' to Y, whereas the base pair Y:X is also allowed.) In Figure 13 are shown a set of secondary structures, linked by single-stranded regions; the conventional nomenclature for the secondary structures includes hairpin loops, asymmetric bulged hairpin loops, symmetric hairpin loops, and pseudoknots.

When nucleotides that are distant in the primary sequence and not thought to interact through Watson/Crick and non-Watson/Crick base pairs are in fact interacting, these interactions (which are often depicted in two dimensions) are also part of the secondary structure.

The three dimensional structure of an RNA motif is merely the description, in space, of the atoms of the RNA motif. Double-stranded RNA, fully base paired through Watson/Crick pairing, has a regular structure in three dimensions, although the exact positions of all the atoms of the helical backbone could depend on the exact sequence of bases in the RNA. A vast literature is concerned with secondary structures of RNA motifs, and those secondary structures containing Watson/Crick base pairs are thought often to form A-form double stranded helices.

From A-form helices one can extend toward the other motifs in three dimensions. Non-Watson/Crick base pairs, hairpin loops, bulges, and pseudoknots are structures built within and upon helices. The construction of these additional motifs is described more fully in the text.

The actual structure of an RNA includes all the atoms of the nucleotide of the molecule in three dimensions. A fully solved structure would include as well bound water and inorganic atoms, although such resolution is rarely achieved by a researcher. Solved RNA structures in three dimensions will include all the secondary structure elements (represented as three dimensional structures) and fixed positions for the

30

atoms of nucleotides not restrained by secondary
structure elements; due to base stacking and other
forces extensive single stranded domains may have fixed
structures.

5   Primary sequences of RNAs limit the possible
three dimensional structures, as do the fixed secondary
structures. The three dimensional structures of an RNA
are limited by the specified contacts between atoms in
two dimensions, and are then further limited by energy
10   minimizations, the capacity of a molecule to rotate all
freely rotatable bonds such that the resultant molecule
is more stable than other conformers having the same
primary and secondary sequence and structure.

          Most importantly, RNA molecules have structures
15   in three dimensions that are comprised of a collection
of RNA motifs, including any number of the motifs shown
in Figure 13.

          Therefore, RNA motifs include all the ways in
which it is possible to describe in general terms the
20   most stable groups of conformations that a nucleic acid
compound can form. For a given target, the ligand
solution and the nucleic acid antibody may be one of
the RNA motifs described herein or some combination of
several RNA motifs.

25   Ligand solutions are defined as the three
dimensional structure held in common or as a family
that define the conserved components identified through
SELEX. For example, the ligands identified for a
particular target may contain a primary sequence in
30   common (NNNCGNAANUCGN'N'N) which can be represented by
a hairpin in two dimensions by:

                         AAN
                     N   U
                     G C
                     C G
35   N N'
                     N N'
                     N N'

40   The three dimensional structure would thus be

SUBSTITUTE SHEET

insensitive to the exact sequence of three of the five
base pairs and two of the five loop nucleotides, and
would in all or most versions of the sequence/structure
be an appropriate ligand for further use. Thus ligand
solutions are meant to represent a potentially large
collection of appropriate sequence/structures, each
identified by the family description which is inclusive
of all exact sequence/structure solutions. It is
further contemplated through this definition that
ligand solutions need not include only members with
exact numerical equivalency between the various
components of an RNA motif. Some ligands may have
loops, for example, of five nucleotides while other
ligands for the same target may contain fewer or more
nucleotides in the equivalent loop and yet be included
in the description of the ligand solution.

Although the ligand solution derived by SELEX
may include a relatively large number of potential
members, the ligand solutions are target specific and,
for the most part, each member of the ligand solution
family can be used as a nucleic acid antibody to the
target. The selection of a specific member from a
family of ligand solutions to be employed as a nucleic
acid antibody can be made as described in the text and
may be influenced by a number of practical
considerations that would be obvious to one of ordinary
skill in the art.

The method of the present invention developed
in connection with investigations of translational
regulation in bacteriophage T4 infection.
Autoregulation of the synthesis of certain viral
proteins, such as the bacteriophage T4 DNA polymerase
(gp43), involves binding of the protein to its own
message, blocking its translation. The SELEX method
was used to elucidate the sequence and structure
requirements of the gp43 RNA binding site. SELEX
allowed the rapid selection of preferred binding

sequences from a population of random nucleic acid
sequences. While exemplified by the isolation and
identification of nucleic acid sequences which bind to
proteins known to bind to RNA, the method of the

5   present invention is generally applicable to the
selection of a nucleic acid capable of binding any
given protein. The method is applicable to selection
of nucleic acids which bind to proteins which do not
(or are not known to) bind to nucleic acid as a part of

10   their natural activity or biological function. The
SELEX method requires no knowledge of the structure or
sequence of a binding site and no knowledge of the
structure or sequence of the target protein. The
method does not depend on purified target protein for

15   selections. In general, application of SELEX will
enrich for ligands of the most abundant target. In a
mixture of ligands, techniques for isolating the ligand
of a given target are available. For example, another
ligand (e.g., substrate, inhibitor, antibody) of the

20   desired target can be used to compete specifically for
binding the target, so that the desired nucleic acid
ligand can be partitioned from ligands of other
targets.

       In the preferred embodiment, ligands derived by
SELEX are comprised of single stranded RNA sequences.

25   It is a critical element of this invention that the
present inventors were able to make conclusions about
RNA that are contrary to those commonly held in the
field, and to use these conclusions to tailor the SELEX

30   process to achieve nucleic acid antibodies derived from
ligand solutions.

       RNA was first appreciated as an information
messenger between the DNA sequences that are the genes
and the protein sequences that are found within enzymes

35   and other proteins. From the first moments after
Watson and Crick described the structure of DNA and the
connection between DNA sequence and protein sequence,

33

the means by which proteins were synthesized became
central to much experimental biochemistry. Eventually
messenger RNA (mRNA) was identified as the chemical
intermediate between genes and proteins. A majority of
5   RNA species present in organisms are mRNAs, and thus
RNA continues to be seen largely as an informational
molecule. RNA serves its role as an informational
molecule largely through the primary sequence of
nucleotides, in the same way that DNA serves its
10   function as the material of genes through the primary
sequence of nucleotides; that is, information in
nucleic acids can be represented in one dimension.
     As the biochemistry of gene expression was
studied, several RNA molecules within cells were
15   discovered whose roles were not informational.
Ribosomes were discovered to be the entities upon which
mRNAs are translated into proteins, and ribosomes were
discovered to contain essential RNA (ribosomal RNAs, or
rRNAs). rRNAs for many years were considered to be
20   structural, a sort of scaffold upon which the protein
components of the ribosome were "hung" so as to allow
the protein components of the ribosome to perform the
protein synthetic action of the ribosome. An
additional large class of RNAs, the transfer RNAs
25   (tRNAs), were postulated and found. tRNAs are the
chemically bifunctional adapters that recognize codons
within mRNA and carry the amino acids that are
condensed into protein. Most importantly, even though
a tRNA structure was determined by X-ray analysis in
30   1974, RNAs were considered to be primarily "strings" in
one dimension for an additional decade. rRNA occupied
a strange position in the research community. For a
long period almost no one sensed the reason behind the
deep similarities in rRNAs from various species, and
35   the true chemical capacity of RNA molecules. Several
researchers postulated that RNA might once have served
an enzymatic rather than informational role, but these

34

postulates were never intended to be predictive about present functions of RNA.

   Tom Cech's work on ribozymes -- a new class of RNA molecules -- expanded the view of the functional capacity of RNA. The group I introns are able to splice autocatalytically, and thus at least some limited catalysis is within the range of RNA. Within this range of catalysis is the activity of the RNA component of RNase P, an activity discovered by Altman and Pace. Cech and Altman received the Nobel Prize in Chemistry for their work, which fundamentally changed the previous limitations for RNA molecules to informational roles. rRNAs, because of the work of Cech and Altman, are now thought by some to be the catalytic center of the ribosome, and are no longer thought to be merely structural.

   It is a central premise of this Invention that RNA molecules remain underestimated by the research community, with respect to binding and other capacities. While ribozymes have caused a remarkable increase in research aimed at RNA functions, the present application contemplates that the shape possibilities for RNA molecules (and probably DNA as well) afford an opportunity to use SELEX to find RNAs with virtually any binding function. It is further contemplated that the range of catalytic functions possible for RNA is broad beyond the present conventional wisdom, although not necessarily as broad as that of proteins.

   The three dimensional shapes of some RNAs are known directly from either X-ray diffraction or NMR methodologies. The existing data set is sparse. The structures of four tRNAs have been solved, as well as three smaller RNA molecules: two small hairpins and a small pseudoknot. The various tRNAs, while related, have elements of unique structure; for example, the anticodon bases of the elongator tRNAs are displayed

SUBSTITUTE SHEET

toward the solvent, while the anticodon bases of an initiator tRNA are pointed more away from the solvent. Some of these differences may result from crystal lattice packing forces, but some are also no doubt a result of idiosyncratic energy minimization by different single stranded sequences within homologous secondary and three dimensional structures.

Sequence variations of course are vast. If a single stranded loop of an RNA hairpin contains eight nucleotides, 65,536 different sequences comprise the saturated sequence "space." Although not bound to the theory of this assertion, the inventors of this invention believe that each member of that set will have, through energy minimization, a most stable structure, and the bulk of those structures will present subtly distinct chemical surfaces to the solvent or to potential interacting target molecules such as proteins. Thus, when all 65,536 sequences within a particular structural motif were tested against the bacteriophage T4 DNA polymerase, two sequences from that set bound better than all others. This suggests that structural aspects of those two sequences are special for that target, and that the remaining 65,534 sequences are not as well suited for binding to the target. It is almost certain that within those 65,536 sequences are other individual members or sets that would be best suited for interacting with other targets.

A key concept in this description of RNA structures is that every sequence will find its most stable structure, even though RNAs are often drawn so as to suggest a random coil or floppy, unstructured element. Homopolymers of RNA, unable to form Watson/Crick base pairs, are often found to have a non-random structure attributed to stacking energy gained by fixing the positions of adjacent bases over each other. Clearly sequences involving all four

SUBSTITUTE SHEET

36

nucleotides may have local regions of fixed structure, and even without Watson/Crick base pairs a non-uniform sequence may have more structure than is at first presumed. The case for fixed structures in RNA loops is even stronger. The anticodon loops of tRNAs have a structure, and so do--presumably--the two winning sequences that bind best to T4 DNA polymerase.

Antiparallel strands of complementary sequence in RNA yield A-form helices, from which loop sequences emerge and return. Even if the loop sequences do not have a strong capacity to interact, energy minimization is an energetically free structure optimization (that is, no obvious energies of activation block energy minimization of a loop sequence). A kinetically likely starting point for optimization may be the loop closing base pair of an RNA stem, which presents a flat surface upon which optimal stacking of loop nucleotides and bases may occur. Loops of RNA are in principle equivalent to loops of protein connecting antiparallel alpha-helices or beta-strands. Although these protein loops are often called random coils, they are neither random nor coiled. Such loops are called "omega" structures, reflecting that the loop emerges and returns to positions that are relatively close to each other (See, Leszczynski, J. and Rose, G. et al. (1986) Science 234:849-855); those positions in a protein are conceptually equivalent to the loop closing base pair of an RNA hairpin.

Many omega structures have been solved by X-ray diffraction, and the structures are idiosyncratic. Clearly each structure is the result of a unique energy minimization acted upon a loop whose ends are close to each other. Both in proteins and RNAs those loops will energy minimize without information from the rest of the structure except, to a first approximation, the loop closing pair of amino acids or base pair. For

both protein omega loops and RNA hairpin loops all the freely rotatable bonds will participate in the attempt to minimize the free energy. RNA, it seems, will be rather more responsive to electrostatics than proteins, while proteins will have many more degrees of freedom than RNAs. Thus, calculations of RNA structures through energy minimization are more likely to yield accurate solution structures than are comparable calculations for proteins.

Single stranded regions of both RNAs and protein may be held so as to extend the possible structure. That is, if a single stranded loop emerges and returns in a protein structure from parallel strands of alpha-helix or beta-strands, the points of emergence and return are further from each other than in the omega structures. Furthermore, the distance spanned by the single strand of peptide can be varied by the lengths of parallel alpha-helix or beta-strand.

For those protein structures in which the single strand lies upon a fixed protein secondary structure, the resultant energy minimization could, in principle, allow interactions between the single stranded domain and the underlying structure. It is likely that amino acid side chains that can form salt bridges in secondary structures could do the same in extended single strands lying on top of regular secondary structures. Thus the exact structures of such protein regions will again be idiosyncratic, and very much sequence dependent. In this case the sequence dependence will include both the single strand and the underlying sequence of the secondary structure.

Interestingly, an RNA structure known as a pseudoknot is analogous to these extended protein motifs, and may serve to display toward solvent or target molecules extended single strands of RNA whose bases are idiosyncratically arrayed toward either the solvent/target or an underlying RNA secondary

structure. Pseudoknots have, in common with protein motifs based on loops between parallel strands, the capacity to alter the length of single strand and the sequence of the helix upon which it lies.

5   Thus, exactly like in protein motifs, by covariation with sequences in the underlying secondary structure it is possible to display single stranded nucleotides and bases toward either the solvent or the underlying structure, thus altering the electrostatics

10   and the functional chemical groups that are interacting with targets. It is important to note that such structure variations follow from energy minimizations, but only one pseudoknot structure is known, even at low resolution. Nevertheless, the value of this Invention

15   arises out of the recognition that the shape and functional displays possible from pseudoknots are recognized to be nearly infinite in unique qualities.

Both hairpin loops and the single stranded domain of pseudoknots are built upon antiparallel RNA

20   helices. Helices of RNA may contain irregularities, called bulges. Bulges can exist in one strand of a helix or both, and will provide idiosyncratic structural features useful for target recognition. Additionally, helix irregularities can provide angled

25   connections between regular helices.

A large bulge (see Figure 13) on one strand of RNA may be comparable to hairpin loops, except that the loop closing base pair is replaced by the two base pairs flanking the bulge.

30   Asymmetric bulges (see Figure 13) may provide an elongated and irregular structure that is stabilized by nucleotide contacts across the bulge. These contacts may involve Watson/Crick interactions or any other stabilizing arrangement, including other hydrogen

35   bonds and base stacking.

Finally, when contemplating fixed RNA shapes or motifs, it is instructive to consider what substantial

differences exist between RNA and proteins. Since
protein is thought to have displaced RNA during
evolution for those activities now carried out almost
entirely by proteins and peptides, including catalysis

5   and highly specific recognition, the chemical
properties of proteins are thought to be more useful
than RNA for constructing variable shapes and
activities. The standard reasoning includes the
existence of 20 amino acids versus only four

10   nucleotides, the strong ionic qualities of lysine,
arginine, aspartic acid, and glutamic acid which have
no counterpart in the RNA bases, the relative
neutrality of the peptide backbone when compared to the
strongly negative sugar-phosphate backbone of nucleic

15   acids, the existence of histidine with a pK near
neutrality, the fact that the side chains of the amino
acids point toward the solvent in both alpha helices
and beta strands, and the regular secondary structures
of proteins. In the double stranded nucleic acids,

20   including RNA, base pairs point the bases toward each
other and utilize much of the chemical information
present at the one dimensional level. Thus, from every
angle presently understood to contribute to shape
diversity and function, proteins are thought to be the

25   vastly superior chemical to nucleic acids, including
RNA. During evolution, proteins were chosen for
recognition and catalysis over RNA, thus supporting the
present widely held view.

      Conversely, and central to this Invention, the

30   vast number of sequences and shapes possible for RNA
will conceivably allow, especially with sequences never
tested during evolutionary history, every desired
function and binding affinity even though RNA is made
up of only four nucleotides and even though the

35   backbone of an RNA is so highly charged. That is, the
RNA motifs described above, with appropriate sequence
specifications, will yield in space those chemical

40

functions needed to provide tight and specific binding
to most targets. It may be suggested that RNA is as
versatile as the immune system. That is, while the
immune system provides a fit to any desired target, RNA

5   provides those same opportunities. The enabling
methodology described herein can utilize 10^{18}
sequences, and thus try vast numbers of structures such
that whatever intrinsic advantages proteins or
specifically antibodies may have over RNA are

10   compensated for by the vastness of the possible "pool"
from which RNA ligands are selected. In addition, with
the use of modified nucleotides, RNA can be used that
is intrinsically more chemically varied than natural
RNAs.

15   The SELEX method involves the combination of a
selection of nucleic acid ligands which bind to a
target molecule, for example a protein, with
amplification of those selected nucleic acids.
Iterative cycling of the selection/amplification steps

20   allows selection of one or a small number of nucleic
acids which bind most strongly to the target from a
pool which contains a very large number of nucleic
acids.

          Cycling of the selection/amplification

25   procedure is continued until a selected goal is
achieved. For example, cycling can be continued until
a desired level of binding of the nucleic acids in the
test mixture is achieved or until a minimum number of
nucleic acid components of the mixture is obtained (in

30   the ultimate case until a single species remains in the
test mixture). In many case, it will be desired to
continue cycling until no further improvement of
binding is achieved. It may be the case that certain
test mixtures of nucleic acids show limited improvement

35   in binding over background levels during cycling of the
selection/amplification. In such cases, the sequence
variation in the test mixture should be increased

SUBSTITUTE SHEET

41

including more of the possible sequence variants or the
length of the sequence randomized region should be
increased until improvements in binding are achieved.
Anchoring protocols and/or walking techniques can be
5   employed as well.

    Specifically, the method requires the initial
preparation of a test mixture of candidate nucleic
acids. The individual test nucleic acids can contain a
randomized region flanked by sequences conserved in all
10   nucleic acids in the mixture. The conserved regions
are provided to facilitate amplification or selected
nucleic acids. Since there are many such sequences
known in the art, the choice of sequence is one which
those of ordinary skill in the art can make, having in
15   mind the desired method of amplification. The
randomized region can have a fully or partially
randomized sequence. Alternatively, this portion of
the nucleic acid can contain subportions that are
randomized, along with subportions which are held
20   constant in all nucleic acid species in the mixture.
For example, sequence regions known to bind, or
selected for binding, to the target protein can be
integrated with randomized regions to achieve improved
binding or improved specificity of binding. Sequence
25   variability in the test mixture can also be introduced
or augmented by generating mutations in the nucleic
acids in the test mixture during the
selection/amplification process. In principle, the
nucleic acids employed in the test mixture can be any
30   length as long as they can be amplified. The method of
the present invention is most practically employed for
selection from a large number of sequence variants.
Thus, it is contemplated that the present method will
preferably be employed to assess binding of nucleic
35   acid sequences ranging in length from about four bases
to any attainable size.

    The randomized portion of the nucleic acids in

the test mixture can be derived in a number of ways.
For example, full or partial sequence randomization can
be readily achieved by direct chemical synthesis of the
nucleic acid (or portions thereof) or by synthesis of a
template from which the nucleic acid (or portions
thereof) can be prepared by use of appropriate enzymes.
End addition, catalyzed by terminal transferase in the
presence of nonlimiting concentrations of all four
nucleotide triphosphates can add a randomized sequence
to a segment. Sequence variability in the test nucleic
acids can also be achieved by employing size-selected
fragments of partially digested (or otherwise cleaved)
preparations of large, natural nucleic acids, such as
genomic DNA preparations or cellular RNA preparations.
In those cases in which randomized sequence is
employed, it is not necessary (or possible from long
randomized segments) that the test mixture contains all
possible variant sequences. It will generally be
preferred that the test mixture contain as large a
number of possible sequence variants as is practical
for selection, to insure that a maximum number of
potential binding sequences are identified. A
randomized sequence of 30 nucleotides will contain a
calculated 10^{18} different candidate sequences. As a
practical matter, it is convenient to sample only about
10^{18} candidates in a single selection. Practical
considerations include the number of templates on the
DNA synthesis column, and the solubility of RNA and the
target in solution. (Of course, there is no
theoretical limit for the number of sequences in the
candidate mixture.) Therefore, candidate mixtures that
have randomized segments longer than 30 contain too
many possible sequences for all to be conveniently
sampled in one selection. It is not necessary to
sample all possible sequences of a candidate mixture to
select a nucleic acid ligand of the invention. It is
basic to the method that the nucleic acids of the test

mixture are capable of being amplified. Thus, it is preferred that any conserved regions employed in the test nucleic acids do not contain sequences which interfere with amplification.

The various RNA motifs described above can almost always be defined by a polynucleotide containing about 30 nucleotides. Because of the physical constraints of the SELEX process, a randomized mixture containing about 30 nucleotides is also about the longest contiguous randomized segment which can be utilized while being able to test substantially all of the potential variants. It is, therefore, a preferred embodiment of this invention when utilizing a candidate mixture with a contiguous randomized region, to use a randomized sequence of at least 15 nucleotides and containing at least about 10^9 nucleic acids, and in the most preferred embodiment contains at least 25 nucleotides.

This invention includes candidate mixtures containing all possible variations of a contiguous randomized segment of at least 15 nucleotides. Each individual member in the candidate mixture may also be comprised of fixed sequences flanking the randomized segment that aid in the amplification of the selected nucleic acid sequences.

Candidate mixtures may also be prepared containing both randomized sequences and fixed sequences wherein the fixed sequences serve a function in addition to the amplification process. In one embodiment of the invention, the fixed sequences in a candidate mixture may be selected in order to enhance the percentage of nucleic acids in the candidate mixture possessing a given nucleic acid motif. For example, the incorporation of the appropriate fixed nucleotides will make it possible to increase the percentage of pseudoknots or hairpin loops in a candidate mixture. A candidate mixture that has been

44

prepared including fixed sequences that enhance the percentage of a given nucleic acid structural motif is, therefore, a part of this invention. One skilled in the art, upon routine inspection of a variety of nucleic antibodies as described herein, will be able to construct, without undue experimentation, such a candidate mixture. Examples 2 and 8 below describe specific examples of candidate mixtures engineered to maximize preferred RNA motifs.

Candidate mixtures containing various fixed sequences or using a purposefully partially randomized sequence may also be employed after a ligand solution or partial ligand solution has been obtained by SELEX. A new SELEX process may then be initiated with a candidate mixture informed by the ligand solution.

Polymerase chain reaction (PCR) is an exemplary method for amplifying of nucleic acids. Descriptions of PCR methods are found, for example in Saiki et al. (1985) Science 230:1350-1354; Saiki et al. (1986) Nature 324:163-166; Scharf et al. (1986) Science 233:1076-1078; Innis et al. (1988) Proc. Natl. Acad. Sci. 85:9436-9440; and in U.S. Patent 4,683,195 (Mullis et al.) and U.S. Patent 4,683,202 (Mullis et al.). In its basic form, PCR amplification involves repeated cycles of replication of a desired single-stranded DNA (or cDNA copy of an RNA) employing specific oligonucleotide primers complementary to the 3' and ·5' ends of the ssDNA, primer extension with a DNA polymerase, and DNA denaturation. Products generated by extension from one primer serve as templates for extension from the other primer. A related amplification method described in PCT published application WO 89/01050 (Burg et al.) requires the presence or introduction of a promoter sequence upstream of the sequence to be amplified, to give a double-stranded intermediate. Multiple RNA copies of the double-stranded promoter containing intermediate

45

are then produced using RNA polymerase. The resultant
RNA copies are treated with reverse transcriptase to
produce additional double-stranded promoter containing
intermediates which can then be subject to another

5   round of amplification with RNA polymerase.
Alternative methods of amplification include among
others cloning of selected DNAs or cDNA copies of
selected RNAs into an appropriate vector and
introduction of that vector into a host organism where

10   the vector and the cloned DNAs are replicated and thus
amplified (Guatelli, J.C. et al. (1990) Proc. Natl.
Acad. Sci. 87:1874). In general, any means that will
allow faithful, efficient amplification of selected
nucleic acid sequences can be employed in the method of

15   the present invention. It is only necessary that the
proportionate representation of sequences after
amplification at least roughly reflects the relative
proportions of sequences in the mixture before
amplification.

20   Specific embodiments of the present invention
for amplifying RNAs were based on Innis et al. (1988)
supra. The RNA molecules and target molecules in the
test mixture were designed to provide, after
amplification and PCR, essential T7 promoter sequences

25   in their 5' portions. Full-length cDNA copies of
selected RNA molecules were made using reverse
transcriptase primed with an ologimer complementary to
the 3' sequences of the selected RNAs. The resultant
cDNAS were amplified by Taq DNA polymerase chain

30   extension, providing the T7 promoter sequences in the
selected DNAs. Double-stranded products of this
amplification proces were then transcribed in vitro.
Transcripts were used in the next
selection/amplification cycle. The method can

35   optionally include appropriate nucleic acid
purification steps.
          In general any protocol which will allow

selection of nucleic acids based on their ability to
bind specifically to another molecule, i.e., a protein
or in the most general case any target molecule, can be
employed in the method of the present invention. It is
only necessary that the selection partition nucleic
acids which are capable of being amplified. For
example, a filter binding selection, as described in
Example 1, in which a test nucleic acid mixture is
incubated with target protein, the nucleic acid/protein
mixture is then filtered through a nitrocellulose
filter and washed with appropriate buffer to remove
free nucleic acids. Protein/nucleic acid often remain
bound to the filter. The relative concentrations of
protein to test nucleic acid in the incubated mixture
influences the strength of binding that is selected
for. When nucleic acid is in excess, competition for
available binding sites occurs and those nucleic acids
which bind most strongly are selected. Conversely,
when an excess of protein is employed, it is expected
that any nucleic acid that binds to the protein will be
selected. The relative concentrations of protein to
nucleic acid employed to achieve the desired selection
will depend on the type of protein, the strength of the
binding interaction and the level of any background
binding that is present. The relative concentrations
needed to achieve the desired selection result can be
readily determined empirically without under
experimentation. Similarly, it may be necessary to
optimize the filter washing procedure to minimize
background binding. Again such optimization of the
filter washing procedures is within the skill of the
ordinary artisan.

A mathematical evaluation of SELEX referred to
as SELEXION has been utilized by the inventors of the
present invention. Appendix A to this application
includes a brief review of the mathematical analysis
utilized to obtain generalizations regarding SELEX

47

derived from SELEXION.

The generalizations obtained from SELEXION are as follows: i) The likelihood of recovering the best-binding RNA in each round of SELEX increases with the number of such molecules present, with their binding advantage versus the bulk RNA pool, and with the total amount of protein used. Although it is not always intuitively obvious to know in advance how to maximize the difference in binding, the likelihood of recovering the best-binding RNA still can be increased by maximizing the number of RNA molecules and target molecules sampled; 2) the ideal nucleic acid and protein concentrations to be used in various rounds of SELEX are dependent on several factors. The experimental parameters suggested by SELEXION parallel those employed in the Examples hereto. For example, when the relative affinity of the ultimate ligand solution is not known--which will almost inevitably be the case when SELEX is performed--it is preferred that the protein and nucleic acid candidate mixture concentrations are selected to provide a binding between about 3 and 7 percent of the total of nucleic acids to the protein target. By using this criterion it can be expected that a tenfold to twentyfold enrichment in high affinity ligands will be achieved in each round of SELEX.

The experimental conditions used to select nucleic acid ligands to various targets in the preferred embodiment are to be selected to mimic the environment that the target would be found _in vivo_. Example 10 below indicates how changing the selection conditions will effect the ligand solution received to a particular target. Although the ligand solution to NGF had significant similarities under high and low salt conditions, differences were observed. Adjustable conditions that may be altered to more accurately reflect the _in vivo_ environment of the target include,

48

but are not limited to, the total ionic strength, the
concentration of bivalent cations and the pH of the
solution. One skilled in the art would be able to
easily select the appropriate separation conditions
5   based on a knowledge of the given target.

In order to proceed to the amplification step,
selected nucleic acids must be released from the target
after partitioning. This process must be done without
chemical degradation of the selected nucleic acids and
10   must result in amplifiable nucleic acids. In a
specific embodiment, selected RNA molecules were eluted
from nitrocellulose filters using a freshly made
solution containing 200 μl of a 7 M urea, 20 mM sodium
citrate (pH 5.0), 1 mM EDTA solution combined with 500
15   μl of phenol (equilibrated with 0.1 M sodium acetate pH
5.2). A solution of 200 μl 7M urea with 500 μl of
phenol has been successfully employed. The eluted
solution of selected RNA was then extracted with ether,
ethanol precipitated and the precipitate was
20   resuspended in water. A number of different buffer
conditions for elution of selected RNA from the filters
can be used. For example, without limitation
nondetergent aqueous protein denaturing agents such as
quanidinium chloride, quanidinium thiocyanate, etc., as
25   are known in the art, can be used. The specific
solution used for elution of nucleic acids from the
filter can be routinely selected by one of ordinary
skill in the art.

Alternative partitioning protocols for
30   separating nucleic acids bound to targets, particularly
proteins, are available to the art. For example,
binding and partitioning can be achieved by passage of
the test nucleic acid mixture through a column which
contains the target molecule bound to a solid support
35   material. Those nucleic acid that bind to the target
will be retained on the column and unbound nucleic
acids can be washed from the column.

Throughout this application, the SELEX process
has been defined as an iterative process wherein
selection and amplification are repeated until a
desired selectivity has been attained. In one
5   embodiment of the invention, the selection process may
be efficient enough to provide a ligand solution after
only one separation step. For example, in theory a
column supporting the target through which the
candidate mixture is introduced--under the proper
10   conditions and with a long enough column--should be
capable of separating nucleic acids based on affinity
to the target sufficiently to obtain a ligand solution.
To the extent that the original selection step is
sufficiently selective to yield a ligand solution after
15   only one step, such a process would also be included
within the scope of this invention.

In one embodiment of this invention, SELEX is
iteratively performed until a single or a discrete
small number of nucleic acid ligands remain in the
20   candidate mixture following amplification. In such
cases, the ligand solution will be represented as a
single nucleic acid sequence, and will not include a
family of sequences having comparable binding
affinities to the target.

25   In an alternate embodiment of the invention,
SELEX iterations are terminated at some point when the
candidate mixture has been enriched in higher binding
affinity nucleic acid ligands, but still contains a
relatively large number of distinct sequences. This
30   point can be determined by one of skill in the art by
periodically analyzing the sequence randomness of the
bulk candidate mixture, or by assaying bulk affinity to
the target.

At this time, SELEX is terminated, and clones
35   are prepared and sequenced. Of course, there will be
an almost unlimited number of clones that could be
sequenced. As seen in the Examples below, however,

after sequencing between 20 and 50 clones it is generally possible to detect the most predominant sequences and defining characteristics of the ligand solution. In a hypothetical example, after cloning 30 sequences it will be found that 6 sequences are identical, while certain sequence portions of 20 of the other sequences are closely related to sequences within the "winning" sequence. Although the most predominant sequence may be considered a ligand solution to that target, it is often more appropriate to construct or describe a ligand solution that consists of a family of sequences that includes the common characteristics of many of the cloned sequences.

In a further embodiment of this invention, a ligand solution that is represented as a family of sequences having a number of defining characteristics (e.g., where the ligand solution is AAGUNNGUNNCNNNN, where N can apparently be any of the four nucleotides) may be used to initiate an additional SELEX process. In this embodiment, the candidate mixture would be comprised of partially fixed and partially random nucleotides, the fixed nucleotides being selected based on the ligand solution received in the initial SELEX process. In this manner, if there is a single nucleotide sequence that binds better than the other members of the ligand solution family, it will be quickly identified.

In an alternate further embodiment of the invention, a second SELEX experiment based on the ligand solution received in a SELEX process is also utilized. In this embodiment, the single most predominant sequence (e.g., AAGUCCGUAACACAC) is used to inform the second SELEX process. In this second SELEX process the candidate mixture is prepared in order to yield sequences based on the selected winner, while assuring that there will be sufficient randomization at each of the sequences. This candidate mixture may be

produced by using nucleotide starting materials that
are biased rather than randomized. For example, the A
solution contains 75% A and 25% U, C and G. Although
the nucleic acid synthesizer is set to yield the
predominant nucleotide, the presence of the other
nucleotides in the A solution will yield nucleic acid
sequences that are predominant in A but that will also
yield variations in this position. Again, this second
SELEX round, informed by the results obtained in the
initial SELEX process, will maximize the probabilities
of obtaining the best ligand solution to a given
target. Again, it must be clarified that the ligand
solution may consist of a single preferred nucleic acid
ligand, or it may consist of a family of structurally
related sequences with essentially similar binding
affinities.

In practice, it may occasionally be preferred
that the SELEX process not be performed until a single
sequence is obtained. The SELEX process contains
several bias points that may affect the predominance of
certain sequences in a candidate mixture after several
rounds of SELEX that are not related to the binding
affinity of that sequence to the target. For example,
a bias for or against certain sequences may occur
during the production of cDNA from the RNA recovered
after selection, or during the amplification process.
The effects of such unpredictable biases can be
minimized by halting SELEX prior to the time that only
one or a small number of sequences predominate in the
reaction mixture.

As stated above, sequence variation in the test
nucleic acid mixture can be achieved or increased by
mutation. For example, a procedure has been described
for efficiently mutagenizing nucleic acid sequences
during PCR amplification (Leung et al. 1989). This
method or functionally equivalent methods can
optionally be combined with amplification procedures in

the present invention.

Alternatively conventional methods of DNA
mutagenesis can be incorporated into the nucleic acid
amplification procedure. Applicable mutagenesis
procedures include, among others, chemically induced
mutagenesis and oligonucleotide site-directed
mutagenesis.

The present invention can also be extended to
utilize additional interesting capacities of nucleic
acids and the manner in which they are known or will
later be found to interact with targets such as
proteins. For example, a SELEX methodology may be
employed to screen for ligands that form Michael
adducts with proteins. Pyrimidines, when they sit in
the correct place within a protein, usually adjacent to
a critical cysteine or other nucleophile, can react
with that nucleophile to form a Michael adduct. The
mechanism by which Michael adducts are formed involves
a nucleophilic attack at the 6 position of the
pyrimidine base to create a transient (but slowly
reversing) intermediate that is really a 5,6
dihydropyrimidine. It is possible to test for the
presence of such intermediates by observing whether
binding between an RNA and a protein target occurs even
after the protein is denatured with any appropriate
denaturant. That is, one searches for a continued
covalent interaction when the binding pocket of the
target has been destroyed. However, Michael adducts
are often reversible, and sometimes so quickly that the
failure to identify a Michael adduct through this test
does not indicate that one was not present at a prior
moment.

SELEX may be done so as to take advantage of
Michael adduct formation in order to create very high
affinity, near-suicide substrates for an enzyme or
other protein target. Imagine that after binding
between a randomized mixture of RNAs and the target,

53

prior to partitioning on a filter or by other means, the target is denatured. Subsequent partitioning, followed by reversal of the Michael adduct and cDNA synthesis on the released RNA, followed by the rest of the SELEX cycle, will enrich for RNAs that bind to a target prior to denaturation but continue to bind covalently until the Michael adduct is reversed by the scientist. This ligand, in vivo, would have the property of permanently inhibiting the target protein. The protein tRNA-uracil methyl transferase (RUMT) binds substrate tRNAs through a Michael adduct. When RUMT is expressed at high levels in E. coli the enzyme is found largely covalently bound to RNA, suggesting strongly that nearly irreversible inhibitors can be found through SELEX.

The method of the present invention has multiple applications. The method can be employed, for example, to assist in the identification and characterization of any protein binding site for DNA or RNA. Such binding sites function in transcriptional or translational regulation of gene expression, for example as binding sites for transcriptional activators or repressors, transcription complexes at promoter sites, replication accessory proteins and DNA polymerases at or near origins of replication and ribosomes and translational repressors at ribosome binding sites. Sequence information of such binding sites can be used to isolate and identify regulatory regions bypassing more labor intensive methods of characterization of such regions. Isolated DNA regulatory regions can be employed, for example, in heterologous constructs to selectively alter gene expression.

It is an important and unexpected aspect of the present invention that the methods described herein can be employed to identify, isolate or produce nucleic acid molecules which will bind specifically to any

54

desired target molecule. Thus, the present methods can
be employed to produce nucleic acids specific for
binding to a particular target. Such a nucleic acid
ligand in a number of ways functionally resembles an
antibody. Nucleic acid ligands which have binding
functions similar to those of antibodies can be
isolated by the methods of the present invention. Such
nucleic acid ligands are designated herein nucleic acid
antibodies and are generally useful in applications in
which polyclonal or monoclonal antibodies have found
application. Nucleic acid antibodies can in general be
substituted for antibodies in any _in vitro_ or _in vivo_
application. It is only necessary that under the
conditions in which the nucleic acid antibody is
employed, the nucleic acid is substantially resistant
to degradation. Applications of nucleic acid
antibodies include the specific, qualitative or
quantitative detection of target molecules from any
source; purification of target molecules based on their
specific binding to the nucleic acid; and various
therapeutic methods which rely on the specific
direction of a toxin or other therapeutic agent to a
specific target site.

   Target molecules are preferably proteins, but
can also include among others carbohydrates,
peptidoglycans and a variety of small molecules. As
with conventional proteinaceous antibodies, nucleic
acid antibodies can be employed to target biological
structures, such as cell surfaces or viruses, through
specific interaction with a molecule that is an
integral part of that biological structure. Nucleic
acid antibodies are advantageous in that they are not
limited by self tolerance, as are conventional
antibodies. Also nucleic acid antibodies do not
require animals or cell cultures for synthesis or
production, since SELEX is a wholly _in vitro_ process.
As is well-known, nucleic acids can bind to

55

complementary nucleic acid sequences. This property of nucleic acids has been extensively utilized for the detection, quantitation and isolation of nucleic acid molecules. Thus, the methods of the present invention are not intended to encompass these well-known binding capabilities between nucleic acids. Specifically, the methods of the present invention related to the use of nucleic acid antibodies are not intended to encompass known binding affinities between nucleic acid molecules. A number of proteins are known to function via binding to nucleic sequences, such as regulatory proteins which bind to nucleic acid operator sequences. The known ability of certain nucleic acid binding proteins to bind to their natural sites, for example, has been employed in the detection, quantitation, isolation and purification of such proteins. The methods of the present invention related to the use of nucleic acid antibodies are not intended to encompass the known binding affinity between nucleic acid binding proteins and nucleic acid sequences to which they are known to bind. However, novel, non-naturally-occurring sequences which bind to the same nucleic acid binding proteins can be developed using SELEX. It should be noted that SELEX allows very rapid determination of nucleic acid sequences that will bind to a protein and, thus, can be readily employed to determine the structure of unknown operator and binding site sequences which sequences can then be employed for applications as described herein. It is believed that the present invention is the first disclosure of the general use of nucleic acid molecules for the detection, quantitation, isolation and purification of proteins which are not known to bind nucleic acids. As will be discussed below, certain nucleic acid antibodies isolatable by SELEX can also be employed to affect the function, for example inhibit, enhance or activate the function, of specific target molecules or

56

structures. Specifically, nucleic acid antibodies can be employed to inhibit, enhance or activate the function of proteins.

Proteins that have a known capacity to bind nucleic acids (such as DNA polymerases, other replicases, and proteins that recognize sites on RNA but do not engage in further catalytic action) yield, via SELEX, high affinity RNA ligands that bind to the active site of the target protein. Thus, in the case of HIV-1 reverse transcriptase the resultant RNA ligand (called 1.1 in Example 2) blocks cDNA synthesis in the presence of a primer DNA, an RNA template, and the four deoxynucleotide triphosphates.

The inventors theory of RNA structures suggests that nearly every protein will serve as a target for SELEX. The initial experiments against non-nucleic acid binding protein were performed with three proteins not thought to interact with nucleic acids in general or RNA in particular. The three proteins were tissue plasminogen activator (tPA), nerve growth factor (NGF), and the extracellular domain of the growth factor receptor (gfR-Xtra). All of these proteins were tested to see if they would retain mixed randomized RNAs on a nitrocellulose filter. tPA and NGF showed affinity for randomized RNA, with Kd's just below uM. gfR-Xtra did not bind with measurable affinity, suggesting that if an RNA antibody exists for that protein it must bind to a site that has no affinity for most other RNAs.

tPA and NGF were taken through the SELEX drill using RNAs with 30 randomized positions. Both tPA and NGF gave ligand solutions in the SELEX drill, suggesting that some site on each protein bound the winning sequences more tightly than that site (or another site) bound other RNAs. The winning sequences are different for the two proteins.

Since tPA and NGF worked so well in the SELEX drill, a random collection of proteins and peptides

SUBSTITUTE SHEET

were tested to see if they had any affinity for RNA.
It was reasoned that if a protein has any affinity for
RNA that the SELEX drill will, on the average, yield
higher affinity sequences which contact the same region
of the target that provides the low, generalized
affinity. A set of proteins and peptides, were tested
to see if randomized RNAs (containing 40 randomized
positions) would be retained on nitrocellulose filters.
About two thirds of the proteins tested bound RNA, and
a few proteins bound RNA very tightly. See Example 9.

Proteins that do not bind RNA to nitrocellulose
filters may fail for trivial reasons having nothing to
do with the likelihood of raising RNA antibodies. One
example, bradykinin, fails to bind to nitrocellulose
filters, and thus would fail in the above experiment.
A bradykinin linked to a solid matrix through the amino
terminus of the peptide was prepared, and then found
that randomized RNA bound tightly to the matrix (see
Example 7). Thus in the initial experiments two short
peptides, bradykinin and bombesin, bind randomized RNAs
quite tightly. Any high affinity RNA ligand obtained
through SELEX with these peptide targets would,
perhaps, be an antagonist of these active peptides, and
might be useful therapeutically. It is difficult to
imagine an RNA of about 30 nucleotides binding to a
very small peptide without rendering that peptide
inactive for virtually any activity.

As described in Examples 4, 7, 9 and 10 below,
proteins not thought to interact with nucleic acids in
nature were found to bind a random mixture of nucleic
acids to a non-trivial extent. It has further been
shown that for such proteins that were found to bind
RNA mixtures non-specifically that a ligand solution
can be obtained following SELEX. It is, therefore, a
potentially valuable screen--prior to the performance
of SELEX--to determine if a given target shows any
binding to a random mixture of nucleic acids.

5

10

15

20

25

30

35

SUBSTITUTE SHEET

58

It is a second important and unexpected aspect
of the present invention that the methods described
herein can be employed to identify, isolate or produce
nucleic acid molecules which will bind specifically to

5   a particular target molecule and affect the function of
that molecule. In this aspect, the target molecules
are again preferably proteins, but can also include,
among others, carbohydrates and various small molecules
to which specific nucleic acid binding can be achieved.

10 Nucleic acid ligands that bind to small molecules can
affect their function by sequestering them or by
preventing them from interacting with their natural
ligands. For example, the activity of an enzyme can be
affected by a nucleic acid ligand that binds the

15 enzyme's substrate. Nucleic acid ligands, i.e.,
nucleic acid antibodies, of small molecules are
particularly useful as reagents for diagnostic tests,
(or other quantitative assays). For example, the
presence of controlled substances, bound metabolites or

20 abnormal quantities of normal metabolites can be
detected and measured using nucleic acid ligands of the
invention. A nucleic acid ligand having catalytic
activity can affect the function of a small molecule by
catalyzing a chemical change in the target. The range

25 of possible catalytic activities is at least as broad
as that displayed by proteins. The strategy of
selecting a ligand for a transition state analog of a
desired reaction is one method by which catalytic
nucleic acid ligands can be selected.

30   It is believed that the present invention for
the first time discloses the general use of nucleic
acid molecules to effect, inhibit or enhance protein
function. The binding selection methods of the present
invention can be readily combined with secondary

35 selection or screening methods for modifying target
molecule function on binding to selected nucleic acids.
The large population of variant nucleic acid sequences

that can be tested by SELEX enhances the probability
that nucleic acid sequences can be found that have a
desired binding capability and function to modify
target molecule activity. The methods of the present
5   invention are useful for selecting nucleic acid ligands
which can selectively affect function of any target
protein including proteins which bind nucleic acids as
part of their natural biological activity and those
which are not known to bind nucleic acid as part of
10   their biological function. The methods described
herein can be employed to isolate or produce nucleic
acid ligands which bind to and modify the function of
any protein which binds a nucleic acid, either DNA or
RNA, either single-stranded or double-stranded; a
15   nucleoside or nucleotide including those having purine
or pyrimidine bases or bases derived therefrom,
specifically including those having adenine, thymine,
guanine, uracil, cytosine and hypoxanthine bases and
derivatives, particularly methylated derivatives,
20   thereof; and coenzyme nucleotides including among
others nicotinamide nucleotides, flavin-adenine
dinucleotides and coenzyme A. It is contemplated that
the method of the present invention can be employed to
identify, isolate or produce nucleic acid molecules
25   which will affect catalytic activity of target enzymes,
i.e., inhibit catalysis or modify substrate binding,
affect the functionality of protein receptors, i.e.,
inhibit binding to receptors or modify the specificity
of binding to receptors; affect the formation of
30   protein multimers, i.e., disrupt quaternary structure
of protein subunits; and modify transport properties of
protein, i.e., disrupt transport of small molecules or
ions by proteins.

The SELEX process is defined herein as the
35   iterative selection and amplification of a candidate
mixture of nucleic acid sequences repeated until a
ligand solution has been obtained. A further step in

SUBSTITUTE SHEET

60

the process is the production of nucleic acid
antibodies to a given target. Even when the ligand
solution derived for a given process is a single
sequence, the nucleic acid antibody containing just the
5   ligand solution must be synthesized. For example, a
SELEX experiment may give a preferred single ligand
solution that consists of only 20 of the 30 randomized
nucleotide sequences used in the SELEX candidate
mixture. The therapeutically valuable nucleic acid
10   antibody would not, preferably, contain the 10 non-
critical nucleotides or the fixed sequences required
for the amplification step of SELEX. Once the desired
structure of the nucleic acid antibody is determined
based on the ligand solution, the actual synthesis of
15   the nucleic acid antibody will be performed according
to a variety of techniques well known in the art.
        The nucleic acid antibody may also be
constructed based on a ligand solution for a given
target that consists of a family of sequences. In such
20   case, routine experimentation will show that a given
sequence is preferred due to circumstances unrelated to
the relative affinity of the ligand solution to the
target. Such considerations would be obvious to one of
ordinary skill in the art.
25   In an alternate embodiment of the present
invention, the nucleic acid antibody may contain a
plurality of nucleic acid ligands to the same target.
For example, SELEX may identify two discrete ligand
solutions. As the two ligand solutions may bind the
30   target at different locations, the nucleic acid
antibody may preferably contain both ligand solutions.
In another embodiment, the nucleic acid antibody may
contain more than one of a single ligand solution.
Such multivalent nucleic acid antibodies will have
35   increased binding affinity to the target unavailable to
an equivalent nucleic acid antibody having only one
ligand.

61

In addition, the nucleic acid antibody may also
contain other elements, that will 1) add independent
affinity for the target to the nucleic acid antibody;
2) dependently enhance the affinity of the nucleic acid
ligand to the target; 3) direct or localize the nucleic
acid antibody to the proper location in vivo where
treatment is desired; or 4) utilize the specifity of
the nucleic acid ligand to the target to effect some
additional reaction at that location.

        The methods of the present invention are useful
for obtaining nucleic acids which will inhibit function
of a target protein, and are particularly useful for
obtaining nucleic acids which inhibit the function of
proteins whose function involves binding to nucleic
acid, nucleotides, nucleosides and derivatives and
analogs thereof. The methods of the present invention
can provide nucleic acid inhibitors, for example, of
polymerases, reverse transcriptases, and other enzymes
in which a nucleic acid, nucleotide or nucleoside is a
substrate or co-factor.

        Secondary selection methods that can be
combined with SELEX include among others selections or
screens for enzyme inhibition, alteration of substrate
binding, loss of functionality, disruption of
structure, etc. Those of ordinary skill in the art are
able to select among various alternatives those
selection or screening methods that are compatible with
the methods described herein.

        It will be readily apparent to those of skill
in the art that in some cases, i.e., for certain target
molecules or for certain applications, it may be
preferred to employ RNA molecules in preference to DNA
molecules as ligands, while in other cases DNA ligands
may be preferred to RNA.

        The selection methods of the present invention
can also be employed to select nucleic acids which bind
specifically to a molecular complex, for example to a

SUBSTITUTE SHEET

substrate/protein or inhibitor/protein complex. Among those nucleic acids that bind specifically to the complex molecules, but not the uncompleted molecules there are nucleic acids which will inhibit the formation of the complex. For example, among those nucleic acids ligands which are selected for specific binding to a substrate/enzyme complex there are nucleic acids which can be readily selected which will inhibit substrate binding to the enzyme and thus inhibit or disrupt catalysis by the enzyme.

An embodiment of the present invention, which is particularly useful for the identification or isolation of nucleic acids which bind to a particular functional or active site in a protein, or other target molecule, employs a molecule known, or selected, for binding to a desired site within the target protein to direct the selection/amplification process to a subset of nucleic acid ligands that bind at or near the desired site within the target molecule. In a simple example, a nucleic acid sequence known to bind to a desired site in a target molecule is incorporated near the randomized region of all nucleic acids being tested for binding. SELEX is then used (Fig. 9) to select those variants, all of which will contain the known binding sequence, which bind most strongly to the target molecule. A longer binding sequence, which is anticipated to either bind more strongly to the target molecule or more specifically to the target can thus be selected. The longer binding sequence can then be introduced near the randomized region of the nucleic acid test mixture and the selection/amplification steps repeated to select an even longer binding sequence. Iteration of these steps (i.e., incorporation of selected sequence into test mixtures followed by selection/amplification for improved or more specific binding) can be repeated until a desired level of binding strength or specificity is achieved. This

63

iterative "walking" procedure allows the selection of
nucleic acids highly specific for a particular target
molecule or site within a target molecule. Another
embodiment of such an iterative "walking" procedure,
5   employs an "anchor" molecule which is not necessarily a
nucleic acid (see Figs. 10 and 11). In this embodiment
a molecule which binds to a desired target, for example
a substrate or inhibitor of a target enzyme, is
chemically modified such that it can be covalently
10   linked to an oligonucleotide of known sequence (the
"guide oligonucleotide" of Fig. 10). The guide
oligonucleotide chemically linked to the "anchor"
molecule that binds to the target also binds to the
target molecule. The sequence complement of guide
15   oligonucleotide is incorporated near the randomized
region of the test nucleic acid mixture. SELEX is then
performed to select for those sequences that bind most
strongly to the target molecule/anchor complex. The
iterative walking procedure can then be employed to
20   select or produce longer and longer nucleic acid
molecules with enhanced strength of binding or
specifity of binding to the target. The use of the
"anchor" procedure is expected to allow more rapid
isolation of nucleic acid ligands that bind at or near
25   a desired site within a target molecule. In
particular, it is expected that the "anchor" method in
combination with iterative "walking" procedures will
result in nucleic acids which are highly specific
inhibitors of protein function (Fig. 11).
30   In certain embodiments of the performance of
SELEX it is desireable to perform plus/minus screening
in conjunction with the selection process to assure
that the selection process is not being skewed by some
factor unrelated to the affinity of the nucleic acid
35   sequences to the target. For example, when selection
is performed by protein binding nitrocellulose, it has
been seen that certain nucleic acid sequences are

64

preferentially retained by nitrocellulose and can be
selected during the SELEX process. These sequences can
be removed from the candidate mixture by incorporating
additional steps wherein the preceding SELEX mixture is
5   passed through nitrocellulose to selectively remove
those sequences selected solely for that property.
Such screening and selection may be performed whenever
the target contains impurities or the selection process
introduces biases unrelated to affinity to the target.

10   SELEX has been demonstrated by application to
the isolation of RNA molecules which bind to and
inhibit the function of bacteriophage T4 DNA
polymerase, also termed gp43. The novel RNA ligand of
T4 DNA polymerase is useful as a specific assay reagent
15   for T4 DNA polymerase. The synthesis of T4 DNA
polymerase is autogenously regulated. In the absence
of functional protein, amber fragments and mutant
proteins are overexpressed when compared to the rate of
synthesis of wild-type protein in replication-deficient
20   infections (Russel (1973) J. Mol. Biol. 79:83-94). In
vitro translation of an N-terminal fragment of gp43 is
specifically repressed by the addition of purified
gp43, and gp43 protects a discrete portion of the mRNA
near its ribosome binding site from nuclease attack
25   (Andrake et al. (1988) Proc. Natl. Acad. Sci. USA
85:7942-7946). The size and sequence of the RNA
translational operator to which gp43 binds and the
strength of that binding have been established. The
minimal size of the gp43 operator is a sequence of
30   about 36 nucleotides, as illustrated in Fig. 1, which
is predicted to have a hairpin loop structure as
indicated therein. The minimal size of the operator
was determined by analysis of binding of end-labeled
hydrolysis fragments of the operator to gp43. Analysis
35   of binding of operator mutants in the hairpin and loop
sequence indicate that gp43 binding to the operator is
sensitive to primary base changes in the helix.

65

Binding to the polymerase was even more reduced by
changes which significantly reduce hairpin stability.
Operator binding was found to be very sensitive to loop
sequence. It was found that replication and operator
5 binding in gp43 are mutually exclusive activities. The
addition of micromolar amounts of purified RNAs
containing intact operator was found to strongly
inhibit in vitro replication by gp43.

   The wild-type gp43 operator, Fig. 1, was
10 employed as the basis for the design of an initial
mixture of RNA molecules containing a randomized
sequence region to assess the ability of the
selection/amplification process to isolate nucleic acid
molecules that bind to a protein. The RNA test mixture
15 was prepared by in vitro transcription from a 110 base
single-stranded DNA template. The template was
constructed as illustrated in Figure 1 to encode most
of the wild-type operator sequence, except for the loop
sequence. The eight base loop sequence was replaced by
20 a randomized sequence region which was synthesized to
be fully random at each base. The template also
contained sequences necessary for efficient
amplification: a sequence at its 3' end complementarily
to a primer for reverse transcription and amplification
25 in polymerase chain reactions and a sequence in its 5'
end required for T7 RNA polymerase transcriptional
initiation and sufficient sequence complementary to the
cDNA of the in vitro transcript. The DNA template is
this a mixture of all loop sequence variants,
30 theoretically containing 65,536 individual species.

   The dissociation constant for the wild-type
loop RNA was found to be about 5×10^{-9}M. The
dissociation constant for the population of loop
sequence variants was measured to be about 2.5×10^{-7}.
35 Randomization of the loop sequence lowered binding
affinity 50-fold.

   In vitro transcripts containing the loop

SUBSTITUTE SHEET

sequence variants were mixed with purified gp43 and incubated. The mixture was filtered through a nitrocellulose filter. Protein-RNA complexes are retained on the filter and unbound RNA is not.

5 Selected RNA was then eluted from the filters as described in Example 1. Selected RNAs were extended with AMV reverse transcriptase in the presence of 3' primer as described in Gauss et al. (1987) supra. The resulting cDNA was amplified with Taq DNA polymerase in

10 the presence of the 5' primer for 30 cycles as described in Innis et al. (1986) supra. The selected amplified DNA served as a template for in vitro transcription to produce selected amplified RNA transcripts which were then subject to another round of

15 binding selection/amplification. The RNA/protein ratio in the binding selection mixture was held constant throughout the cycles of selection. The iterative selection/amplification was performed using several different RNA/protein molar ratios. In all experiments

20 RNA was in excess: experiment A employed an RNA/gp43 of 10/1 (moles/moles); experiment B employed an RNA/gp43 of 1000/1; and experiment C employed an RNA/gp43 of 100/1.

The progress of the selection process was
25 monitored by filter binding assays of labelled transcripts of amplified cDNA at the completion of each cycle of the procedure. Batch sequencing of the RNA products from each round for experiment B was also done to monitor the progress of the selection.

30 Autoradiograms of sequencing gels of RNA products after 2, 3 and 4 rounds of selection/amplification are shown in Figure 3. It is clear that there was no apparent loop sequence bias introduced until after the third selection. After the fourth round of selection, an

35 apparent consensus sequence for the eight base loop sequence is discernable as: A(a/g)(u/c)AAC(u/c)(u/c). Batch sequencing of selected RNA after the fourth round

SUBSTITUTE SHEET

67

of selection for experiments A, B and C is compared in Figure 4. All three independent SELEX procedures using different RNA/protein ratios gave similar apparent consensus sequences. There was, however, some apparent

5   bias for wild-type loop sequence (AAUAACUC) in the selected RNA from experiments A and C.

In order to determine what allowable sequence combinations were actually present in the selected RNAs, individual DNAs were cloned from selected RNAs

10   after the fourth round of selection in experiment B. The batch sequence result from experiment B appeared to indicate an even distribution of the two allowable nucleotides which composed each of the four variable positions of the loop sequence. Individuals were

15   cloned into pUC18 as described by Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, (Cold Spring Harbor, N.Y.), Sections 1.13; 1.85-1.86. Twenty individual clones that were identified by colony filter hybridization to the 3' primer were sequenced. None of

20   the sequenced clones were mutant at any place in the operator sequence outside of the loop sequence. Only five variant sequences were observed as shown in Figure 7, and surprisingly only two sequence variants were the major components of the selected mixture. The

25   frequencies of each sequence in the 20 individual isolates sequenced are also given in Figure 7. The wild-type sequence AAUAACUC and the loop AGCAACCU were present in approximately equal amount in the selected RNA of experiment B. The other selected variants were

30   1 base mutants of the two major variants. The strength of binding of the sequence variants was compared in filter binding assays using labelled in vitro transcripts derived from each of the purified clonal isolates. As shown in Figure 6, a rough correlation

35   between binding affinity of an RNA for gp43 and the abundance of the selected sequence was observed. The two major loop sequence variants showed approximately

68

equal binding affinities for gp43.

The loop sequence variant RNAs isolated by the selection/amplification process, shown in Figure 7, can all act as inhibitors of gp43 polymerase activity as has been demonstrated for the wild-type operator sequence.

An example of the use of SELEX has been provided by selection of a novel RNA ligand of bacteriophage T4 DNA polymerase (gp43) (Andrake et al. (1988) Proc. Natl. Acad. Sci. USA 85:7942-7946).

The present invention includes specific ligand solutions, derived via the SELEX process, that are shown to have an increased affinity to HIV-1 reverse transcriptase, R17 coat protein, HIV-1 rev protein, HSV DNA polymerase, E. coli ribosomal protein S1, tPA and NGF. These ligand solutions can be utilized by one of skill in the art to synthesize nucleic acid antibodies to the various targets.

The following examples describe the successful application of SELEX to a wide variety of targets. The targets may generally be divided into two categories--those that are nucleic acid binding proteins and those proteins not known to interact with nucleic acids. In each case a ligand solution is obtained. In some cases it is possible to represent the ligand solution as a nucleic acid motif such as a hairpin loop, an asymmetric bulge or a pseudoknot. In other examples the ligand solution is presented as a primary sequence. In such cases it is not meant to be implied that the ligand solution does not contain a definitive tertiary structure.

In addition to T4 DNA polymerase, targets on which SELEX has been successfully performed include bacteriophage R17 coat protein, HIV reverse transcriptase (HIV-RT), HIV-1 rev protein, HSV DNA polymerase plus or minus cofactor, E. coli ribosomal protein S1, tPA and NGF. The following experiments

also describe a protocol for testing the bulk binding
affinity of a randomized nucleic acid candidate mixture
to a variety of proteins. Example 7 also describes the
immobilization of bradykinin and the results of bulk

5   randomized nucleic acid binding studies on bradykinin.

        The examples and illustrations herein are not
to be taken as limiting in any way. The fundamental
insight underlying the present invention is that
nucleic acids as chemical compounds can form a

10   virtually limitless variety of sizes, shapes and
configurations and are capable of an enormous
repertoire of binding and catalytic functions, of which
those known to exist in biological systems are merely a
glimpse.

15

        ### EXAMPLES

        The following materials and methods were used
throughout.

        The transcription vector pT7-2 is commercially

20   available (U.S. Biochemical Company, Cleveland, OH).
Plasmid pUC18 is described by Norrander et al. (1983)
Gene 24:15-27 and is also commercially available from
New England Biolabs. All manipulations of DNA to
create new recombinant plasmids were as described in

25   Maniatis et al. (1982) Molecular Cloning: A Laboratory
Manual, Cold Spring Harbor Laboratory, Cold Spring
Harbor, New York, except as otherwise noted. DNA
olignucleotides were synthesized and purified as
described in Gauss et al. (1987) Mol. Gen. Genet.

30   206:24-34.

        In vitro transcriptions with T7 RNA polymerase
and RNA gel-purification were performed as described in
Milligan et al. (1987) Nucl. Acids Res. 15:8783-8798,
except that in labeling reactions the concentrations of

35   ATP, CTP, and GTP were 0.5 mM each, and the UTP
concentration was 0.05 mM. The UTP was labeled at the
alpha position with ^{32}P at a specific activity of

approximately 20 Ci/mmol. Crude mRNA preparations from
T4 infections, labeling of oligos, and primer extension
with AMV reverse transcriptase were all according to
Gauss et al. (1987) supra.

Dilutions of labeled, gel-purified RNA and
purified gp43 were made in 200 mM potassium acetate, 50
mM Tris-HCl pH 7.7 at 4°C. In nitrocellulose filter
binding assays, purified gp43 was serially diluted and
30 μl aliquots of each dilution of protein were added
to 30 μl aliquots of diluted, labeled, gel-purified
RNA. The RNA dilution (50 μl) was spotted on a fresh
nitrocellulose filter, dried and counted to determine
input counts per tube. The concentration of protein in
the reactions ranged from 10^{-10} M to 10^{-8} M and the
concentration of the RNAs in each experiment was
approximately 10^{-12} M. After incubation at 4°C for 30
minutes, each tube was placed at 37°C for 3 minutes and
50 μl of each sample filtered through pre-wet
nitrocellulose filters (Millipore #HAWP 025 00) and
washed with 3 ml of 200 mM potassium acetate, 50 mM
Tris-HCl pH 7.7. The filters were dried and counted in
Ecolume™ scintillation fluid (ICN Biomedicals, Inc.).
Controls were done in the absence of gp43, from which
the background (always less than about 5% of the input
counts) was determined. From each set of measurements
the background was subtracted, and the percent of total
input counts remaining on the filters calculated. From
each set of data points, a best-fit theoretical
bimolecular binding curve was generated using a version
of a published program (Caceci and Cacheris, 1984
supra) modified to construct a curve described by the
equation,

$$\Theta = A[gp43]/(Kd + [gp43])$$

where Θ is the fraction of the total RNA that is bound
to the filter, A is the percent of RNA at which binding
saturates (approximately 60% for this protein-RNA
interaction), [gp43] is the input gp43 concentration,

71

and Kd is the dissociation constant for the bimolecular reaction. This equation is an algebraic rearrangement of equation [1-5] from Bisswanger (1979) Theorie und Methoden der Enzymkinetik, Verlag Chemie, Weinheim, FRG, p. 9 with the simplifying assumption that the concentration of the protein far exceeds the concentration of RNA-protein complexes, an assumption which is valid in the experiments described.

Example 1.   Selection of RNA Inhibitors of T4 DNA Polymerase

A 110 base single-stranded DNA template for in vitro transcription was created as shown in Figure 2 by ligation of three synthetic oligonucleotides (Tables 1, 3, 4 and 5) in the presence of two capping oligonucleotides (Tables 1 and 2). One of the template-creating oligos was also used as the 3' primer in reverse transcription of the in vitro transcript and subsequent amplification in polymerase chain reactions (PCRs) (Innis et al. (1988) Proc. Natl. Acad. Sci. USA 85:9436-9440). One of the capping oligos (1) contains the information required for T7 RNA polymerase transcriptional initiation and sufficient sequence complementarily to the cDNA of the in vitro transcript to serve as the 5' primer in the PCR amplification steps. The DNA template encoded an RNA which contains the entire RNA recognition site for T4 DNA polymerase except that a completely random sequence was substituted in place of the sequence which would encode the wild-type loop sequence AAUAACUC. The random sequence was introduced by conventional chemical synthesis using a commercial DNA synthesizer (Applied Biosystems) except that all four dNTP's were present in equimolar amounts in the reaction mixture for each position indicated by N in the sequence of oligonucleotide number 4 (Table 1). The random sequence is flanked by primer annealing sequence

information for the 5' and 3' oligos used in PCR. The
DNA template is thus a mixture of all loop sequence
variants, theoretically containing 65,536 individual
species. The dissociation constant for the wild-type
loop variant RNA sequence is about 5×10^{-9} M and for
the population of sequences was measured to be about
2.5×10^{-7} M, a 50-fold lower binding affinity.

TABLE 1

1)   5'-TAATACGACTCACTATAGGGAGCCAACACCACAATTCCAATCAAG-3'

2)   5'-GGGCTATAAACTAAGGAATATCTATGAAAG-3'

3)   5'-GAATTGTGGTGTTGGCTCCCTATAGTGAGTCGTATTA-3'

4)   5'-ATATTCCTTAGTTTATAGCCCNNNNNNNNNNAGGCTCTTGATTG-3'
and

5)   5'-GTTTCAATAGAGATATAAAATTCTTTCATAG-3'

In *vitro* transcripts containing the loop
sequence variants were mixed with purified gp43 at
three different RNA-protein ratios throughout the
multiple rounds of selection. (For A and B the
concentration of gp43 was 3×10^{-8} M, "low protein,"
and for C the concentration of gp43 was 3×10^{-7} M,
"high protein." For A the concentration of RNA was
about 3×10^{-7}, "low RNA," and for B and C the
concentration of RNA was about 3×10^{-5} M, "high RNA.")
One round consisted of the following steps:

1)   Selection. The RNA and protein were mixed
in the desired ratios described above, incubated at
37°C, washed through a nitrocellulose filter, and RNA
was eluted from the filters as described *supra*.

2)   Amplification. The RNA eluted from filters
was extended with AMV reverse transcriptase in the

presence of 50 picomoles of 3' promer in a 50 μl reaction under conditions described in Gauss et al. (1987) supra. To the resulting cDNA synthesis 50 picomoles of 5' primer was added and in a reaction volume of 100 μl and was amplified with Taq DNA polymerase as described in Innis (1988) supra for 30 cycles.

3) Transcription. In vitro transcription is performed on the selected amplified templates as described in Milligan et al. (1987) supra, after which DNaseI is added to remove the DNA template. The resultant selected RNA transcripts were then used in step 1 of the next round. Only one-twentieth of the products created at each step of the cycle were used in the subsequent cycles so that the history of the selection could be traced. The progress of the selection method was monitored by filter binding assays of labeled transcripts from each PCR reaction. After the fourth round of selection and amplification, the labeled selected RNA products produced binding to gp43 equivalent to that of wild-type control RNA. The RNA products from each round for one experiment (B) and from the fourth round for all three experiments were gel-purified and sequenced. In Figure 3, we show the sequence of the purified in vitro transcripts derived from the second, third and fourth rounds of selection and amplification for experiment B. It is clear that there was no apparent loop sequence bias introduced until after the third selection. By this point in the selection, there was a detectable bias which was complete by the fourth round for the apparent consensus sequence A(a/g)(u/c)AAC(u/c)(u/c). Batch sequencing of the RNA transcribed after the fourth selection and amplification for trials A, B, and C is shown in Figure 4. All three independent runs with different protein/RNA ratios gave similar results. There is some

74

apparent bias for wild-type sequence at each of the
four "variable" positions in experiments A and C.

In order to find out what allowable
combinations actually existed, we used two "cloning"
oligonucleotides which contained restriction site
information, to amplify sequences from RNA from the
fourth round of experiment B from which individuals
were cloned into pUC18 as described (Sambrook et al.
(1989) supra; Innis et al. (1988) supra). The selected
batches of trial B were chosen for further examination
because there appeared to be an even distribution of
the two allowable nucleotides which composed each of
the four "variable" positions. Twenty individual
clones that were identified by colony filter
hybridization to the 3' primer were sequenced. None of
these individuals were mutant at any place in the
operator sequence outside of the loop sequence
positions that were deliberately varied. The sequence
distributions are summed up in Figure 7. Surprisingly,
the selected RNA mixture was actually composed of two
major loop sequences. One was the wild-type sequence,
AAUAACUC of which 9 out of 20 were isolated. The
other, AGCAACCU, was mutant at four positions and
existed in 8 of the 20 clones (see Figure 7). The
other three loop sequences detected were single
mutations of these two major sequences. Filter binding
experiments with labeled in vitro transcripts derived
from each of these clonal isolates indicated that there
was a rough correlation between binding affinity of an
RNA for gp43 and selected abundance (see Figure 7).

Example 2:   Isolation of a specific RNA ligand for HIV
              reverse transcriptase

The reverse transcriptase activity of HIV-1 is
composed of a heterodimer of two subunits (p51 and p66)
that have common amino termini. The extra
carboxyterminal region of the larger peptide comprises

the RNaseH domain of reverse transcriptase; the
structure of that domain has recently been determined
at high resolution.

It has been previously shown that this HIV-1
reverse transcriptase directly and specifically
interacts with its cognate primer tRNA^{Lys3} to which it
was experimentally cross-linked at the anti-codon loop
and stem. It was also found that only the heterodimer
exhibited this specific RNA recognition; neither
homodimeric species of reverse transcriptase bound with
specificity to this tRNA.

Two template populations (with approximately
10^{14} different sequences each) were created for use in
SELEX by ligation. One template population was
randomized over 32 nucleotide positions, using fixed
sequences at the ends of the randomized region to
afford cDNA synthesis and PCR amplification. The
second template population had, as additional fixed
sequence at the 5' end of the RNA, the anticodon loop
and stem of tRNA^{Lys3}. (All oligos used in this work are
shown in Table 2). There was no difference in the
affinity of the two randomized populations for HIV-1
reverse transcriptase [RT] (and, as is shown, the RNAs
which were selected did not utilize either 5' region in
specific binding). Nine rounds of SELEX with each
population were performed using the heterodimer HIV-RT
as the target protein.

The mechanism by which the randomized DNA was
prepared utilizing ligations and bridging
oligonucleotides was described previously. Such
methodology can diminish the total number of different
sequences in the starting population from the
theoretical limit imposed by DNA synthesis at the 1
micromole scale.

In these ligation reactions about 1 nanomole of
each oligonucleotide was used. The ligated product was
gel-purified with an approximate yield of 50%. This

purified template was transcribed with T7 RNA
polymerase as described above. It was found that HIV
RT could saturably bind this random population with a
half-maximal binding occuring at about 7×10^{-7} M as
5   determined by nitrocellulose assays. All RNA-protein
binding reactions were done in a binding buffer of 200
mM KOAc, 50 mM Tris-HCl pH 7.7, 10 mM dithiothreitol.
RNA and protein dilutions were mixed and stored on ice
for 30 minutes then transferred to 37°C for 5 minutes.
10   (In binding assays the reaction volume is 60 μl of
which 50 ul is assayed; in SELEX rounds the reaction
volume is 100 ul). Each reaction is suctioned through
a prewet (with binding buffer) nitrocellulose filter
and rinsed with 3 mls of binding buffer after which it
15   is dried and counted for assays or subjected to elution
as part of the SELEX protocol. Nine rounds were
performed. The RNA concentration for all nine rounds
was approximately 3×10^{-5} M. HIV-RT was 2×10^{-8} M in
the first selection and 1×10^{-8} M in selections 2-9.
20   The experiment using RNA containing the tRNA^{Lys3}
anticodon loop and stem was completed first.
Nitrocellulose filter binding assays performed at the
ninth round revealed that the RNA population had
increased about 100-fold in affinity to HIV-1 RT when
25   compared to the starting canidate mixture, but that the
background binding to nitrocellulose filters in the
absence of protein had increased from about 2% of input
RNA to 15%. Individual sequences were cloned from this
population (after filtration through nitrocellulose
30   filters to delete some of the high background of
potential sequences selected for retention by filters
alone) and are listed in Table 3. Nitrocellulose
filter binding assays of selected sequences' affinity
for HIV RT are shown in Figure 14. Some of the
35   sequences were selected as ligands for HIV-RT,
exemplified by the binding curves of ligands 1.1 and
1.3a, and show some sequence homology as illustrated by

Tables 4 and 5. Some of the ligand sequences exhibit significant retention on nitrocellulose filters in the absence of protein, exemplified by ligand 1.4 (Figure 14), and seem to be characterized by a long helix with a loop of purine repeat elements (as shown in Table 4). In spite of our minimal, late efforts to delete them in this experiment prior to cloning, these sequences represented a significant part of those collected from this experiment.

As a consequence, experiment 2 (which has a different 5' fixed sequence) was pre-filtered through nitrocellulose before the first, third, sixth and ninth rounds of selection. The sequences collected from this experiment are shown in Table 6. There are again many sequences with homology to those of high affinity from experiment 1 as shown in Tables 4 and 5. There are many fewer, if any, sequences that fit the motif of sequences retained by nitrocellulose filters alone. Nitrocellulose binding assays of selected ligand sequences from this experiment compared to that of ligand 1.1 are shown in Figure 15.

High affinity ligand RNAs with the most common sequence (1.1) and a similar sequence (1-3a) were further analyzed to determine the boundaries of the information required for high affinity binding to HIV-1 RT. The results of these experiments are shown in Figure 16. These experiments establish that the motif common to these sequences, UUCCGNNNNNNNNCGGGAAA, are similarly positioned within the recognition domain. The sequences UUCCG and CGGGA of this motif may base-pair to form an RNA helix with an eight base loop. In order to discover what besides these fixed sequences may contribute to high affinity binding to HIV-1 RT, a canidate mixture template was created that contained random incorporation at the nucleotide positions that differ from these two sequences as shown in Table 7. After eight rounds of SELEX, individual sequences were

78

cloned and sequenced. The 46 sequences are shown in
Table 7. Inspection of these sequences reveals
extensive base-pairing between the central 8n variable
region and the downstream 4n variable region and
5   flanking sequences; base-pairing which in combination
with that discussed above would indicate an RNA
pseudoknot. That no specific sequences predominate in
this evolved population suggests that there is no
selection at the primary sequence level and that
10   selection occurs purely on the basis of secondary
structure, that is, there are many sequence
combinations that give similar affinities for HIV-1 RT,
and none have competitive advantage. Analysis of the
first and second SELEX experiments reveals that the
15   individual sequences which comprise those populations
that have homology to the UUCCG...CGGGANAA motif also
show a strong potential for this pseudoknot base-
pairing.

Figure 31 shows a schematic diagram of what is
20   referred to herein as a pseudoknot. A pseudoknot is
comprised of two helical sections and three loop
sections. Not all pseudoknots contain all three loops.
For the purposes of interpreting the data obtained, the
various sections of the pseudoknot have been labeled as
25   shown in Figure 31. For example, in Table 5 several of
the sequences obtained in experiments one and two are
listed according to the pseudoknot configuration
assumed by the various sequences.

The results of experiments one and two, as
30   defined in Table 5, led to experiment three wherein
sequences in S1(a), S1(b) and L3 were fixed. Again,
the SELEX derived nucleic acids were configured almost
exclusively in pseudoknots. Examination of the results
in each of the experiments reveals that the nucleic
35   acid solution to HIV-RT contains a relatively large
number of members, the most basic common denominator
being that they are all configured as pseudoknots.

79

Other generalizations defining the nucleic acid
solution for HIV-RT are as follows:

    1)   S1(a) often comprises the sequence
5'-UUCCG-3' and S1(b) often comprises the sequence
5'-CGGGA-3'. However, base pair flips are allowed, and
the stem may be shortened.

    2)   L1 may be short or long, but often
comprises two nucleotides in the best binding nucleic
acids. The 5' nucleotide in L1 often is either a U or
an A.

    3)   S2 is usually comprised of 5 or 6 base
pairs, and appears to be sequence independent. This
stem may contain non-Watson/Crick pairs.

    4)   L2 may be comprised of no nucleotides, but
when it exists, the nucleotides are preferably A's.

    5)   L3 is generally 3 or more nucleotides,
enriched in A.

    6)   In most sequences obtained by SELEX, the
total number of nucleotides in L1, S2(a) and L2 equals
8.

A primary purpose of this experiment was to
find ligand solutions to HIV-1 RT. The ability of the
evolved ligand clone 1.1 was compared to the ability of
the starting population for experiment 1 to inhibit
reverse transcriptase activity, and is shown in Figure
17. Even at equal concentrations of inhibitor RNA to
RT, the reverse transcriptase is significantly
inhibited by ligand 1.1. In contrast, only at 10 mM
(or 200-fold excess) starting population RNA is there
any significant inhibition of the HIV-1 RT. Thus, the
high affinity ligand to HIV-1 RT either blocks or
directly interacts with the catalytic site of the
enzyme.

In order to test the specifity of this
inhibition, various concentrations of ligand 1.1 were
assayed for inhibition of MMLV, AMV and HIV-1 reverse
transcriptase. The results of that experiment which

SUBSTITUTE SHEET

are shown in Figure 18 show that the inhibition of
ligand 1.1 is specific to HIV-1 reverse transcriptase.

Example 3:   Isolation of specific RNA ligand for
                bacteriophage R17 coat protein.

5

       SELEX was performed on the bacteriophage R17
coat protein. The protein was purified as described by
Carey et al., Biochemistry, 22, 2601 (1983). The
binding buffer was 100mM potassium acetate plus 10 mM

10   dithiothreitol plus 50 mM Tris-acetate pH 7.5. Protein
and RNA were incubated together for three minutes at
37°C and then filtered on nitrocellulose filters to
separate protein-bound RNA from free RNA. The filters
were washed with 50 mM Tris-acetate pH 7.5. Protein

15   was at 1.2 x
10^{-7}M for the first four rounds of SELEX and at 4 x 10^{-8}
for rounds five through 11.

       The starting RNA was transcribed from DNA as
described previously. The DNA sequence includes a

20   bacteriophage T7 RNA polymerase promoter sequence that
allows RNA to be synthesized according to standard
techniques. cDNA synthesis during the amplification
portion of the SELEX cycle is primed by a DNA of the
sequence:

25   cDNA primer (PCR primer 1):

       5'GTTTCAATAGAGATATAAAATTCTTTCATAG 3'

       The DNA primers used to amplify the cDNA was,
thus, the sequence including the T7 promoter, 32
randomized positions, an AT dinucleotide, and the fixed

30   sequence complementary to PCR primer 1. The RNA that
is used to begin the first cycle of SELEX thus has the
sequence:

pppGGGAGCCAACACCACAAUUCCAAUCAAG-32N-AUCUAUGAAAGAAUUUUAUCUCUAUUGAAAC

       A set of clones from after the 11th round of

35   SELEX was obtained and sequenced. Within the 38
different sequences obtained in the 47 clones were
three found more than once: one sequence was found six

times, one sequence four times, and another two times.
The remaining 35 sequences were found once each. Two
sequences were not similar to the others with respect
to primary sequences or likely secondary structures,
and were not analyzed further. Thirty-six sequences
had in common the sequence ANCA situated as a
tetranucleotide loop of a bulged hairpin; the bulged
nucleotide was an adenine in all 36 cases. The
sequences of the entire set are given in Table 8,
aligned by the four nucleotides of the hairpin loop.
The two nucleotides 3' to the randomized portion of the
starting RNA (an AU) are free to change or be deleted
since the cDNA primer does not include the
complementary two nucleotides; many clones have changed
one or both of those nucleotides.

     The winning RNA motif, shown in Figure 19,
bears a direct relationship to the coat binding site
identified earlier through site-directed mutagenesis
and binding studies. <u>See</u>, Uhlenbeck <u>et al</u>. <u>supra</u>
(1983); Ramaniuk <u>et al</u>. <u>supra</u> (1987). However, some of
the sequences are more conserved in this set than might
have been expected. The loop sequence AUCA
predominates, while earlier binding data might have
suggested that ANCA sequences are all equivalent. The
natural binding site on the R17 genome includes the
sequence and structure shown below:

```
              UU
         A         A
           GC
           GC
         A
           GC
```
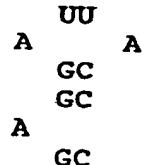
     The natural structure includes the sequence
GGAG, which serves to facilitate ribosome binding and
initiation of translation of the R17 replicase coding
region. During SELEX that requirement is not present,
and the winning sequences contain around the loop and
bulge C:G base pairs more often than G:C base pairs.
SELEX, therfore, relaxes the constraints of biology and

82

evolutionary history, leading to ligands with higher
affinities than the natural ligand. Similarly, the
loop cytidine found in each of the 36 sequences is a
uridine in the natural site, and it is known that C
provides higher affinity than U. During evolution
natural sites must have an appropriate affinity rather
than the highest affinity, since the tightest binding
may lead to disadvantages for the organism.

Example 4:   Isolation of a nucleic acid ligand for a
              serine protease.

Serine proteases are protein enzymes that
cleave peptide bonds within proteins. The serine
proteases are members of a gene family in mammals, and
are important enzymes in the life of mammals. Serine
proteases are not known to bind to nucleic acids.
Examples of serine proteases are tissue plasminogen
activator, trypsin, elastase, chymotrypsin, thrombin,
and plasmin. Many disease states can be treated with
nucleic acid ligands that bind to serine proteases, for
example, disorders of blood clotting and thrombus
formation. Proteases other than serine proteases are
also important in mammalian biology, and these too
would be targets for nucleic acid ligands with
appropriate affinities obtained according to the
invention herein taught.

Human tissue plasminogen activator (htPA),
available from commercial sources, was chosen as a
serine protease to place through the SELEX method of
this invention. The RNA candidate mixture used was
identical to that described in Example 11 below in the
HSV DNA polymerase experiment.

Binding during SELEX was in 50 mM NaCl plus 50
mM Tris-acetate pH 7.5 for 3 minutes at 37 degrees.
SELEX was carried out for ten rounds. The 30N
candidate mixture bound to tPA with an affinity (kd) of
$7 \times 10(-8)$ M in 150 mM NaAc plus 50 mM Tris-acetate pH

7.5; the affinity of the RNA present after nine rounds
of SELEX was about threefold tighter. Nine clones were
isolated, sequenced, and some of these were tested for
binding to tPA as pure RNAs. The sequences of the nine
5   clones obtained at low salt were as follows:

Name	#	Sequence of random region
A1	3	ACGAAACAAAUA<u>AGGAGGAGGAGG</u>GAUUGU
A2	1	<u>AGGAGGAGGAGG</u>GAGAGCGCAAAUGAGAUU
A3	1	<u>AGGAGGAGGAGG</u>UAGAGCAUGUAUUAAGAG
B	1	UAAGCAAGAAUCUACGAUAAAUACGUGAAC
C	1	AGUGAAAGACGACAACGAAAAACGACCACA
D	1	CCGAGCAUGAGCCUAGUAAGUGGUGGAUA
E	1	UAAUAAGAGAUACGACAGAAUACGACAUAA

All tested sequences bound at least somewhat
better than the starting 30N candidate mixture.
However, the A series bound to nitocellulose better in
the absence of tPA than did the candidate mixture, as
20   though the shared sequence motif caused retention on
the nitrocellulose matrix by itself. That motif is
underlined in the sequences shown above. In other
SELEX experiments AGG repeats have been isolated when
trying to identify a ligand solution to HIV-1 reverse
25   transcriptase, the human growth hormone receptor
extracellular domain, and even the R17 coat protein in
a first walking experiment. When tested, these
sequences show modest or substantial binding to
nitrocellulose filters without the target protein being
30   present. It appears that the AGG repeats may be found
in hairpin loops. Since SELEX is an iterative process
in most embodiments, it is not surprising that such
binding motifs would emerge.

The existence of nitrocellulose binding motifs
35   may be avoided by one or more of several obvious
strategies. RNA may be filtered through the
nitrocellulose filters prior to SELEX to eliminate such

SUBSTITUTE SHEET

84

motifs. Alternative matrices may be used in
alternative rounds of SELEX, e.g., glass fiber filters.
Alternative partitioning systems may be used, e.g.,
columns, sucrose gradients, etc. It is obvious that
5   any given single process will lead to biases in the
iterative process that will favor motifs that do not
have increased binding to the target, but are selected
by the selection process. It is, therefore, important
to use alternating processes or screening processes to
10   eliminate these motifs. It has been shown that the AGG
repeats, like other motifs isolated as biases that are
target independent, will tend to emerge most frequently
when the affinity of the best sequences for the target
are rather low or when the affinities of the best
15   sequences are only slightly better than the affinity of
the starting candidate mixture for the target.

Example 5:   Isolation of a nucleic acid ligand for a
mammalian receptor
20   Mammalian receptors often are proteins that
reside within the cytomplasmic membranes of cells and
respond to molecules circulating outside of those
cells. Most receptors are not known to bind to nucleic
acids. The human growth hormone receptor responds to
25   circulating human growth hormone, while the insulin
receptor responds to circulating insulin. Receptors
often have a globular portion of the molecule on the
extracellular side of the membrane, and said globular
portion specifically binds to the hormone (which is the
30   natural ligand). Many disease states can be treated
with nucleic acid ligands that bind to receptors.
Ligands that bind to a soluble globular domain
of the human growth hormone receptor (shGHR) are
identified and purified using the candidate mixture of
35   Example 4. Again, the binding buffers are free of DTT.
The soluable globular domain of the human growth
hormone receptor is available from commercial and

SUBSTITUTE SHEET

academic sources, having usually been created through recombinant DNA technology applied to the entire gene encoding a membrane-bound receptor protein. SELEX is used reiteratively until ligands are found. The ligands are cloned and sequenced, and binding affinities for the soluble receptor are measured. Binding affinities are measured for the same ligand for other soluble receptors in order to ascertain specifity, even though most receptors do not show strong protein homologies with the extracellular domains of other receptors. The ligands are used to measure inhibition of the normal binding activity of shGHR by measuring competitive binding between the nucleic acid ligand and the natural (hormone) ligand.

Example 6:   Isolation of a nucleic acid ligand for a mammalian hormone or factor

Mammalian hormones or factors are proteins, e.g., growth hormone, or small molecules (e.g., epinephrine, thyroid hormone) that circulate within the animal, exerting their effects by combining with receptors that reside within the cytoplasmic membranes of cells. For example, the human growth hormone stimulates cells by first interacting with the human growth hormone receptor, while insulin stimulates cells by fist interacting with the insulin receptor. Many growth factors, e.g., granulocyte colony stimulating factor (GCSF), including some that are cell-type specific, first interact with receptors on the target cells. Hormones and factors, then, are natural ligands for some receptors. Hormones and factors are not known, usually, to bind to nucleic acids. Many disease states, for example, hyperthyroidism, chronic hypoglycemia, can be treated with nucleic acid ligands that bind to hormones or factors.

Ligands that bind to human insulin are identified purified using the starting material of

86

Example 3.   Human insulin is available from commercial
sources, having usually been created through
recombinant DNA technology.   SELEX is used
reiteratively until a ligand is found.   The ligands are
cloned and sequenced, and the binding affinities for
human insulin are measured.   Binding affinities are
measured for the same ligand for other hormones or
factors in order to ascertain specificity, even though
most hormones and factors do not show strong protein
homologies with human insulin.   However, some hormone
and factor gene families exist, including a small
family of IGF, or insulin-like growth factors.   The
nucleic acid ligands are used to measure inhibition of
the normal binding activity of human insulin to its
receptor by measuring competitive binding with the
insulin receptor and the nucleic acid ligand in the
presence or absence of human insulin, the natural
ligand.

<u>Example 7</u>:   <u>Preparation of column matrix for SELEX</u>
        Following the procedures as described in
Example 9 below, it was shown that the polypeptide
bradykinin is not retained by nitrocellulose.   To
enable the SELEX process on bradykinin, the protein was
attached to Activated CH Sepharose 4B (Pharmacia LKB)
as a support matrix according to standard procedures.
The resulting matrix was determined to be 2.0 mM
bradykinin by ninhydrin assay.   See Crestfield <u>et al</u>.
<u>J. Biol. Chem</u>. vol. 238, pp. 238, pp. 622-627 (1963);
Rosen <u>Arch. Biochem. Biophys</u>., vol. 67, pp. 10-15
(1957).   The activated groups remaining on the support
matrix were blocked with Tris.   See Pharmacia, <u>Affinity
Chromatography: Principles and Methods</u>, Ljungforetagen
AB, Uppsala, Sweden (1988).
        Spin-column separation was used to contact
solutions of candidate mixtures with beaded matrix.   In
a general procedure for performing a selection step for

SELEX, 40 µL of a 50:50 slurry of target sepharose in
reaction buffer is transferred to a 0.5 ml Eppendorf
tube. The RNA candidate mixture is added with 60 µL of
reaction buffer, the reaction mixture is allowed to
equilibrate for 30 minutes at 37°C. A hole is pierced
in the bottom of the tube, and the tube is placed
inside a larger Eppendorf tube, both caps removed, and
the tubes spun (1000 RPM, 10", 21°C) to separate the
eluate. The small tube is then transferred to a new
larger tube, and the contents washed four times by
layering with 50 µL of the selected wash buffer and
spinning. To conduct binding assays, the tube
containing the radioactive RNA is transferred to a new
Eppendorf tube and spun to dryness.

A bulk binding experiment was performed wherein
a RNA candidate mixture comprised of a 30 nucleic acid
randomized segment was applied to the bradykinin
sepharose matrix. Using the spin-column technique, the
binding of the bulk 30N RNA to various matrices was
determined under high salt concentrations to determine
the best conditions for minimizing background binding
to the sepharose. Background binding of RNA to
sepharose was minimized by blocking activiated groups
on the sepharose with Tris, and using a binding buffer
of 10 mM DEM and 10-20 mM KOAc. At this buffer
condition, a binding curve of the randomized bulk
solution of RNA yielded a bulk Kd of about 1.0×10^{-5}.
See Figure 20. The curve was determined by diluting
the bradykinin sepharose against blocked, activated
sepharose.

Example 8:   Preparation of candidate mixtures enhanced
              in RNA motif structures.

In the preferred embodiment, the candidate
mixture to be used in SELEX is comprised of a
contiguous region of between 20 and 50 randomized
nucleic acids. The randomized segment is flanked by

88

fixed sequences that enable the amplification of the
selected nucleic acids.

In an alternate embodiment, the candidate
mixtures are created to enhance the percentage of
nucleic acids in the candidate mixture possessing given
nucleic acid motifs. Although two specific examples
are given here, this invention is not so limited. One
skilled in the art would be capable of creating
equivalent candidate mixtures to achieve the same
general result.

In one specific example, shown as Sequence A in
Figure 21, the candidate mixture is prepared so that
most of the nucleic acids in the candidate mixture will
be biased to form a helical region of between 4 and 8
base pairs, and a "loop" of either 20 or 21 contiguous
randomized sequences. Both 5' and 3' ends of the
sequence mixture will contain fixed sequences that are
essential for the amplification of the nucleic acids.
Adjacent these functional fixed sequences will be fixed
sequences chosen to base pair with fixed sequences on
the alternate side of the randomized region. Going
from the 5' to the 3' end of the sequences, there will
be 5 distinct regions: 1) fixed sequences for
amplification; 2) fixed sequences for forming a helical
structure; 3) 20 or 21 randomized nucleic acid
residues; 4) fixed sequences for forming a helical
structure with the region 2 sequences; and 5) fixed
sequences for amplification. The A candidate mixture
of Figure 21 will be enriched in hairpin loop and
symmetric and asymmetric bulged motifs. In a preferred
embodiment, the candidate mixture would contain equal
amounts of sequences where the randomized region is 20
and 21 bases long.

A second example, shown in Figure 21 as
sequence B, is designed to enrich the candidate mixture
in nucleic acids held in the psuedoknot motif. In this
candidate mixture, the fixed amplification sequences

89

flank three regions of 12 randomized positions. The
three randomized regions are separated by two fixed
regions of four nucleotides, the fixed sequences
selected to preferably form a four basepair helical
structure. Going from the 5' to the 3' end of the
sequence, there will be 7 district regions: 1) fixed
sequences for amplification; 2) 12 randomized
nucleotides; 3) fixed sequences for forming a helical
structure; 4) 12 randomized nucleotides; 5) fixed
sequences for forming a helical structure with the
region 3 nucleotides; 6) 12 randomized nucleotides; and
7) fixed sequences for amplification.

In a preferred candidate mixture, the
engineered helical regions are designed to yield
alternating GC, CG, GC, CG basepairs. This basepair
motif has been shown to give a particularly stable
helical structure.

Example 9:   Bulk binding of randomized RNA sequences
              to proteins not known to bind nucleic
              acids.

Following the general nitrocellulose selection
procedures as described in Example 1 above for SELEX, a
group of randomly selected proteins were tested to
determine if they showed any affinity to a bulk
candidate mixture of RNA sequences. The candidate
mixture utilized in each experiment consisted of a 40N
RNA solution (a randomized mixture having a 40
randomized nucleic acid segment) that was radiolabled
to detect the percentage of binding. The candidate
mixture was diluted in binding buffer (200 mM KoAc, 50
mM TrisoAc pH 7.7, 10 mM DTT) and 30 μL was used in a
60 μL binding reaction. To each reaction was added 20
μL, 10 μL or 1 μL of each protein. Binding buffer was
added to reach a total volume of 60 μL. The reactions
were incubated at 37°C for 5 minutes and then subjected
to filter binding.

90

The proteins tested were Acetylcholinesterase (MW 230,000); N-acetyl-β-D-glucosaminidase (MW 180,000); Actin (MW 43,000); Alcohol Dehydrogenase (240,000); Aldehyde Dehydrogenase (MW 200,000);

5   Angiotensin (MW 1297); Ascorbate Oxidase (MW 140,000); Atrial Nutriuretic Factor (MW 3,064); and Bombesin (MW 1621). The proteins were purchased from Boehringer Ingelheim, and were utilized in the buffer composition in which they are sold.

10   The RNA candidate mixture used in each experiment contained 10,726 counts of radiolable, and a background binding of about 72 counts was found. The results are summarized in Table 9. All proteins tested except Acetylcholinesterase, N-acetyl-β-D-

15   glucosaminidase and Actin were found to yield some bulk RNA affinity. Because of the low concentration of N-acetyl-β-D-glucosaminidase in solution as purchased, the results for that protein are not definitive. In addition, if any of the proteins tested do not bind to

20   nitrocellulose--which is the case for bradykinin--no affinity would be detected in this experiment. Example 7 above discussing column supported bradykinin demonstrates that the failure to show bulk binding in this experiment does not mean that bulk binding does

25   not exist for a given protein.

Example 10: Isolation of RNA ligand solution for Nerve Growth Factor.

Nerve growth factor (NGF) is a protein factor

30   that acts through a receptor on the outside surfaces of target cells. Antagonists toward growth factors and other hormones can act by blocking a receptor or by titrating the factor or hormone. An RNA was sought by the SELEX process that binds directly to NGF.

35   The starting RNAs were prepared exactly as in the case of HSV DNA polymerase (Example 11).

Two different experiments were done with NGF.

SUBSTITUTE SHEET

The first was a ten round SELEX using low salt binding
buffer, 3 minutes at 37 degrees incubation, and then
filtration and a wash with the same buffer during the
SELEX. The low salt binding buffer was 50mM NaCl plus
5   50 mM Tris-acetate pH 7.5. The second experiment used
as the binding buffer 200 mM NaCl plus 50 mM Tris-
acetate pH 7.5, and then after filtration a wash with
50 mM Tris-acetate pH 7.5; this SELEX experiment went
through only seven rounds.

10   The low salt experiment yielded 36 cloned
sequences. Fifteen of the clones were nearly identical
– #'s 2, 3, 4, 5, 6, 8, 11, 13, 19, 22, 28, 33, and 34
were identical, while #'s 15 and 25 had a single
difference:

15   ACAUCGAUGACCGGAAUGCCGCACACAGAG
          +A   G
        (15) (25)
            A second abundant sequence, found six times,
was:

20   CCUCAGAGCGCAAGAGUCGAACGAAUACAG (#'s 12, 20, 27, and 31)

        From the high salt SELEX ten clones have been
sequenced, but eight of them are identical and
obviously related to the abundant (but minor) second
25   class from the low salt experiment. The winning
sequence is:
            ----CUCAUGGAGCGCAAGACGAAUAGCUACAUA----
            Between the two experiments a total of 14
different sequences were obtained (sequences with one
30   difference are lumped together in this analysis); they
are listed here, with the similarities overmarked and
the frequencies noted. ngf.a through ngf.k are from
the low salt experiment, while hsngf.a through hsngf.c
are from the high salt experiment:
35

                        xxxxxxxxxx   ######   Frequency
        ngf.a   ACAUCGAUGACCGGAAUGCCGCACACAGAG   15/36

                        xxxxxxxxxx   ######
        ngf.b   CCUCAGAGCGCAAGAGUCGAACGAAUACAG   6/36
                $$$$$$$$$$$$$$ $$$$ $$$$

92

```
            #######    xxxxxxxxxx
ngf.c       CGGGUGAUUAGUACUGCAGAGCGGAAUGAC              5/36

            #######    xxxxxxxxxx
ngf.d       UGCGAUAAGACUUGCUGGGCGAGACAAACA              3/36

            #######        xxxxxxxxxx
ngf.e       AGUCCACAGGGCACUCCCAAAGGGCAAACA              1/36

               xxxxxxxxxx#######
ngf.f       ACUCAUGGAGCGCUCGACGAUCACCAUCGA              1/36

            xxxxxxxxxx        #######
ngf.g       CAAGGGAGAAUGUCUGUAGCGUCCAGGUA               1/36

            xxxxxxxxxx    #######
ngf.h       CGACGCAGAGAUACGAAUACGACAAAACGC              1/36

            #######xxxxxxxxxx
ngf.i       GAGGGUGAGGUGGGAGGCAGUGGCAGUUUA              1/36

               xxxxxxxxxx#######
ngf.j       UGAACUAGGGGGGAGGGAGGGUGGAAGACAG             1/36

               #######  xxxxxxxxxx
ngf.k       GUGGAGGGGACGUGGAGGGGAGAGCAAGA               1/36

               xxxxxxxxxx#######
hsngf.a     CUCAUGGAGCGCAAGACGAAUAGCUACAUA             8/10
            $$$$  $$$$$$$$$$$$$     $$$$

            xxxxxxxxxx    #######
hsngf.b     CCAUAGAGGCCACAAGCAAAGACUACGCA              1/10

            #######    xxxxxxxxxx
hsngf.c     CCUACAAGAAAAGAGGGAAGGAGAAAAAAA            1/10
```

While no obvious secondary structure is embedded within the similar sequences, it is likely that the winning sequences place critical nucleotides into a structure that is well fit by an NGF binding site.

A binding assay of nucleic acid hsngf.a to NGF was performed, and this nucleic acid was found to have a Kd of about 20 to 30 fold higher than the bulk 30N candidate mixture. The same nucleic acid was also found to have a lower or equal affinity to R17 coat protein and tPA than a 30N candidate mixture. Thus,

93

the SELEX derived nucleic acid ligand hsngf.a is a selective ligand to NGF.

Example 11: Isolation of a nucleic acid ligand for HSV-1 DNA polymerase.

Herpes simplex virus (HSV-1) is a DNA-containing virus of mammals. HSV-1, like many DNA-containing viruses, encodes its own DNA polymerase. The HSV-1 DNA polymerase has been purified in two forms, which have different qualities but each of which will catalyze DNA replication in vitro. The simple form, which is one polypeptide, is purified from cells expressing the cloned gene according to Hernandez, T.R. and Lehman, I.R., J. Biol. Chem., 265, 11227-11232 (1990). The second form of DNA polymerase, a heterodimer, is purified from HSV-1 infected cells according to Crute, J.J. and Lehman, I.R., J. Biol. Chem., 264, 19266-19270 (1989); the heterodimer contains one peptide corresponding to the polymerase itself and another, UL42, also encoded by HSV-1.

SELEX was performed on both the single polypeptide and the heterodimer. The binding buffer in each case was 50 mM potassium acetate plus 50 mM Tris acetate, pH 7.5, and 1 mM dithiothreitol. Filtration to separate bound RNA was done after four minutes of incubation at 37 degrees; the filters were washed with binding buffer minus dithiothreitol.

The RNA candidate mixture was transcribed from DNA as described previously. As is the case in other embodiments, the DNA sequence includes a bacteriophage T7 RNA polymerase promoter sequence that allows RNA to be synthesized according to standard techniques. cDNA synthesis during the amplification portion of SELEX is primed by a DNA of the sequence:

cDNA primer (PCR primer 1): 5' GCCGGATCCGGGCCTCATGTGAA 3'

94

The DNA primers used to amplify the cDNA in
that portion of the SELEX cycle include, in one of
them, the T7 promoter; that PCR primer has the
sequence:

5PCR primer 2: 5' CCGAAGCTTAATACGACTCACTATAGGGAGCTCAGAATAAACGCTCAA
3'

The initial randomized DNA included the
sequence with the T7 promoter, 30 randomized positions,
and the fixed sequence complementary to PCR primer 1.

10   The RNA that is used to begin the first cycle of SELEX
thus has the sequence:

pppGGGAGCUCAGAAUAAACGCUCAA - 30N - UUCGACAUGAGGCCCGGAUCCGGC

SELEX was performed for seven rounds, after
which cDNA was prepared and cloned as described

15   previously. The series of sequences designated "H"
were obtained with the simple HSV DNA polymerase as the
target, while the "U" series was obtained with the
heterodimeric polymerase that includes the UL42
polypeptide.

20   About 25% of the sequences from the H series
contain an exact sequence of 12 nucleotides at the 5'
end of the randomized region (the upper case letters
are from the randomized region). In some sequences the
length between the fixed primers was not exactly 30

25   nucleotides, and in one case (H2) a large deletion was
found within the randomized region. The members of
this H subset include:

```
                    XXXXXXXXXXXX
        H5:    --cgcucaaUAAGGAGGCCACGGACAACAUGGUACAGCuucgaca--
30      H10:   --cgcucaaUAAGGAGGCCACAACAAAIGGAGACAAAuucgaca--
        H4:    --cgcucaaUAAGGAGGCCACACACAUAGGUAGACAUGuucgaca--
        H19:   --
        cgcucaaUAAGGAGGCCACAUACAAAAGGAUGAGUAAAuucgaca--
        H20:   --
35      cgcucaaUAAGGAGGCCACAAAUGCUGGUCCACCGAGAuucgaca--
        H38:   --cgcucaaUAGGGAGGGCACGGGAAGGGUGAGUGGAUAuucgaca-
        -

        H2:    --cgcucaaUAAGGAGGCCACAAGuucgaca--
```

95

Two members of the U series share this primary
sequence motif:

U9:   --cgcucaaUAAGGAGGGCCACAGAUGUAAUGGAAACuucgaca--
U13:   --
cgcucaaUAAGGAGGCCACAUACAAAAGGAUGAGUAAAAuucgaca--

The remaining sequences from the H and U series
show no obvious common sequence; in addition, no
sequences from the seventh round emerged as winning
single sequences in either series, suggesting that more
rounds of SELEX will be required to find the best
ligand family for inhibiting HSV DNA polymerase.

It appears that the primary sequence.
--cgcucaaUAAGGAGGCCAC . . . .
may be a candidate for an antagonist species, but those
members of the series have yet to be tested as
inhibitors of DNA synthesis. It appears that the fixed
sequence just 5' to the UAAGGAGGCCAC must participate
in the emergence of this subset, or the shared 12
nucleotides would have been positioned variably within
the randomized region.

Example 12:   Isolation of a nucleic acid ligand
for   E. coli Ribosomal Protein S1:

The E. coli 30S ribosomal protein S1 is the
largest of the 21 30S proteins. The protein has been
purified based on its high affinity for
polypyrimidines, and is thought to bind rather tightly
to single stranded polynucleotides that are pyrimidine
rich. It was questioned if the RNA identified as a
ligand solution by SELEX was in any way more
information rich than a simple single stranded RNA rich
in pyrimidines.

The RNAs, DNAs, cDNA primer (PCR primer 1), and
PCR primer 2 were identical to those used for HSV-1 DNA
polymerase (see, Example 11) . The binding buffer
contained 100 mM ammonium chloride plus 10 mM magnesium
chloride plus 2 mM dithiothreitol plus 10 mM Tris-

chloride, pH 7.5.   Binding was at room temperature, and
complexes were once again separated by nitrocellulose
filtration.   The protein was purified according to I.
Boni et al., European J. Biochem., <u>121</u>, 371 (1982).

5   After 13 SELEX rounds, a set of 25 sequences
was obtained.   More than twenty of those sequences
contained pseudoknots, and those pseudoknots contain
elements in common.

              The general structure of pseudoknots can be
10   diagrammed as;

              STEM 1a - LOOP 1 - STEM 2a - STEM 1b - LOOP 2 -
STEM 2b (<u>See</u> Figure 31)

              Most of the S1 protein ligands contain:

              STEM 1 of 4 to 5 base pairs, with a G just 5'
15   to LOOP 1

              LOOP 1 of about 3 nucleotides, often ACA

              STEM 2 of 6 to 7 base pairs, stacked directly
upon STEM   1

              LOOP 2 of 5 to 7 nucleotides, often ending with
20   GGAAC

              A reasonable interpretation of these data is
that LOOP 2 is stretched across STEM 1 so as to hold
that loop rigidly in a form that simplifies and
enhances the binding of the single strand to the active
25   site of protein S1.   A picture of the consensus
pseudoknot in two dimensions would look like this:

```
                              ┌──── R N G
                             ╱┌─── Y          G
30                          ╱╱┌── (C/G)         A
                           ╱╱╱┌─ (U/A)         A
                          ╱╱╱ ╱   C   C
                         ╱╱╱ ╱    N-N'
                        ╱╱╱ ╱     N-N'
                       ╱╱╱ ╱      A-u
35                     ╱╱╱ ╱      A-u
                      │││ │       G-c
       5'--NNNYR(G/C)(A/U)GACAC-gNNNNNNNN--3'
```
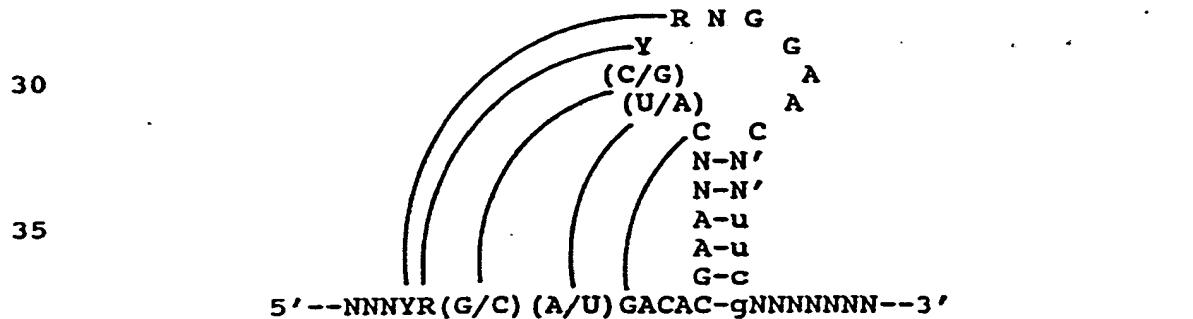
              In such figures the base pairs are shown as
40   lines and dashes, the selections of bases from the
randomized region are shown in upper case letters, Y is

97

a pyrimidine, R is a purine, N- N' means any base pair, N means any nucleotide, and the lower case letters are from the fixed sequence used for PCR amplifications.

It appears that single stranded polynucleotide binding proteins and domains within proteins will often select, during SELEX, a pseudoknot which presents the extended, rigid single strand called LOOP 2 to the binding site of the protein in a manner that maximizes the interactions with that site. Thus, when the HIV-1 RT psueodoknot emerged, it is reasonable to think that the single stranded domain LOOP 2 is bound within the region of RT that holds the template strand during replication. That is, it appears reasonable that most replication enzymes (DNA polymerase, RNA polymerase, RNA replicases, reverse transcriptases) will have a domain for holding the template strand that might prefer a pseudoknot as the ligand of choice from SELEX.

Example 13:   Isolation of a nucleic acid ligand to HIV-1 rev protein

The HIV-1 rev protein's RNA-recognition site appears to be complex, and its function is essential to the productive infection of an epidemic viral disease. See, Olsen et al., Science, vol. 247, pp. 845-848 (1990). The SELEX process on this protein was performed in order to learn more about the recognition element and to isolate a ligand to the target protein.

A candidate mixture was created with a 32 nucleotide long random region as described above in Example 2. It was found that the rev protein could saturably bind the starting candidate mixture with a half-maximal binding occuring at about $1 \times 10^{(-7)}$ M as determined by nitocellulose assays. All RNA-protein binding reactions were performed in a binding buffer of 200 mM KOAc, 50 mM Tris-HCl pH 7.7, 10 mM dithiothreitol. RNA and protein dilutions were mixed and stored on ice for 30 minutes then transferred to 37

98

degrees for 5 minutes.   (In binding assays the reaction
volume is 60 ul of which 50 ul is assayed; in SELEX
rounds the reaction volume is 100 ul.) Each reaction
is suctioned through a prewet (with binding buffer)

5   nitrocellulose filter and rinsed with 3 mls of binding
buffer after which it is dried and counted for assays
or subjected to elution as part of the SELEX protocol.
Ten rounds of SELEX were performed, using a RNA
concentration of about 3 x 10(-5) M. The concentration

10   of rev protein was 1 x 10(-7) in the first round, and
2.5 x 10(-8) in all subsequent rounds. The intial
candidate mixture was run over a nitrocellulose filter
to reduce the number of sequences that have a high
affinity for nitrocellulose. This process was also

15   repeated after rounds 3, 6, and 9. The cDNA product
was purified after every third round of selection to
avoid anomalously sized species which will typically
arise with repeated rounds of SELEX. After 10 rounds
the sequence in the variable region of the RNA

20   population was nonrandom as determined by dideoxy-chain
termination sequencing. 53 isolates were cloned and
sequenced.

          Each of the cloned sequences are listed in
Table 10. All sequences were analysed by the Zucker

25   RNA secondary structure prediction program. See,
Zucker, Science, vol. 244, pp. 48-52 (1989); Jaeger et
al., Proc. Natl. Acad. Sci. USA, vol. 86, pp. 7706-7710
(1989). On the basis of common secondary structure all
sequences have been grouped into three common motifs as

30   shown in Table 11. Motifs I and II are similiar in
conformation including a bulged loop closed at each end
by a helix. This generalized structure has been
illustrated schematically in Table 12, and the domains
labeled for easy discussion; that is from 5' to 3' Stem

35   la (which base pairs to the 3' Stem 1b), Loop 1, Stem
2a, Loop 3, Stem 2b, Loop 2, and Stem 1b. The
sequences which fit in the various domains are listed

for individual sequences in Table 12. (Note that in sequence 3a, the homologous alignment is flipped 180 degrees so that it is Stem 1 which is closed with a loop.) The energies of folding of the RNA molecule (including the fixed flanking sequences) are shown in Table 13.

The wild-type rev responsive element (RRE) that has been determined to be at least minimally involved in binding of rev to HIV-1 transcripts was also folded by this program, and is included in Tables 12 and 13.

The sequences were also searched for related subsequences by a procedure based on that described in Hertz et al. Comput. Appl. Biosci., vol.6. pp.81-92 (1990). Two significant patterns were identified. Each isolate was scored to identify its best match to the patterns, the results of which can be seen in Table 13. The related subsequences motifs are presented by the common secondary structures in similiar conformations; that is, the first sequence UUGAGAUACA is commonly found as Loop 1 plus the 3' terminal CA, which pairs with the UG at the 5' end of the second information rich sequence UGGACUC (commonly Loop 3). There is also a strong prediction of base-pairing of the GAG of sequence I to the CUC of sequence II. Motif II is similiar to Motif I in that the subsequence GAUACAG predominates as a loop opposite CUGGACAC with a similiar pairing of CA to UG. Motif II differs in the size of the loops and some of the sequence particularly in the absence of predicted base-pairing across the loop. One domain of the wild-type RRE closely resembles Motif II. Motif III is the least like all the other sequences, although it is characterized by two bulged U's adjacent to base-paired GA-UC as in Motif I. Unfortunately, further comparisons are complicated because the folding pattern of Motif III involves the 3' fixed sequence region in critical secondary structures; because these sequences are

100

invariant there is no way to analyse the importance of
any one of them. The folded sequences of
representatives of each Motif is shown in Figure 23
with the folded sequence of the wild-type RRE.

5   The sequences were further analyzed for their
affinity to the rev protein. Templates were PCR'd from
a number of clones from which labeled in vitro
transcripts were prepared and individually assayed for
their ability to bind rev protein. These binding

10 curves are shown in Figures 24 to 28. Labeled
transcripts from oligonucleotide templates were also
synthesized which contain the wild-type RRE discussed
above, and what is inferred to be the consensus motif
in a highly stable conformation. To control for

15 experimental variations, the best binding sequence,
isolate 6a, was assayed as a standard in every binding
experiment. The RNA-protein mixtures were treated as
described above except that diluted RNA's were heated
to 90 degrees for 1 minute and cooled on ice prior to

20 mixing. The average Ka for isolate 6a was 8.5 x 10(-8)
M, and the results of this experiment are shown in
Table 13.

          The binding curves of Figure 24 shows that the
evolved population (P) improved approximately 30fold

25 for binding to rev protein relative to the starting
candidate mixture. The binding of the wild-type RRE
closely resembles that of the most abundant clone, 1c.
This experiment also illustrates how sensitive the rev
binding interaction is to secondary structure.

30 Isolates 6a and 6b are identical in the regions of high
information content, but are quite different at the
level of secondary structure resulting in changes at
three nucleotide positions. These changes, which
predict the base-pairing of Stem 1, lower the affinity

35 of 6b by 24fold. Sensitivity to secondary structure
anomalies is further illustrated by the binding of
isolate 17 as shown in Figure 25. Isolate 17 has the

101

maximum information score as shown in Table 12.
However, there is an extra bulged U at the 5' end of
Loop 1 as shown in Table 11. This extra U results in
isolate 17's reduced affinity for rev as compared to
5   other sequences of Motif I. In contrast, single
nucleotide deletions of Loop 2 sequences, even those
that diminish the prospect of cross-bulge base-pairing
are well tolerated by the rev interaction.

         Another compelling commonality is the
10   conservation of the sequence ACA opposite UGG where the
CA pairs with the UG to begin Stem 2. This sequence is
shared by Motifs I and II as well as by the wild-type
RRE. Sequences 11 and 12 exhibit a base-pair
substitution at this position (see Table 12), and
15   sequence 12 was tested and has reduced affinity
compared to most of the other Motif I sequences.

         The RNA sequences determined by SELEX to be rev
ligands may be classified by primary and secondary
structure. A consensus emerges of an asymmetric bulge
20   flanked by two helices in which are configured
specifically conserved single and double stranded
nucleotides. Although base-pairing across the bulge is
predicted for many of the sequences isolated (Motif I),
it may not be essential or crucial to rev interaction.
25   Optimal sizes for Loop 1 appear to be 8 (Motif I) or 6
(Motif III) where there is an observed penalty for
sizes of 9 or 3. Optimal sizes for Loop 3 are 5 and 4.
In addition, the interaction of rev with the various
domains of these ligands may be additive. Motif II
30   resembles Moftif I primarily at the junction of Loops 1
and 3 at Stem 2. Motif III resembles Motif I at the
junction of Loops 1 and 3 at Stem 1. Consensus
diagrams of the Motif I and II nucleic acid solutions
for HIV-rev are shown in Figures 29 and 30.

35   The abundance of sequences in the cloned
population is not strictly correlated with affinity to
rev protein. It is possible that the concentration of

rev protein used throughout the SELEX process was sufficient to bind a signifigant percentage of all these isolates. As a consequence, there may have been selection for replicability of cDNA and DNA during PCR superimposed on a low stringency selection for binding to rev. The highly structured nature of these ligands and the possible differences in the efficiency of cDNA synthesis on these templates reinforces this potential replicative bias. Also, there is some mutation that occurs during the SELEX process. The sequence 6a so resembles 6b that they must have a common ancestor. This relatively late arrival during the rounds of SELEX may explain the paucity of this sequence irrespective of its higher affinity to the target. In the same manner, some of the ligands that have emerged may have mutated relatively recently during selection from ancestor sequences that exist in the intial candidate mixture but are not represented in the cloned population.

The invention disclosed herein is not limited in scope to the embodiments disclosed herein. As disclosed, the invention can be applied by those of ordinary skill in the art to a large number of nucleic acid ligands and targets. Appropriate modifications, adaptations and expedients for applying the teachings herein in individual cases can be employed and understood by those skilled in the art, within the scope of the invention as disclosed and claimed herein.

TABLE 2

1a) 5'-taatacgactcactatagggagccaacaccacaattccaatcaag-3'
     (bridging oligo for 5' construction and 5'PCR oligo)

1b) 5'-taatacgactcactatagggagcatcagactttttaatctgacaatcaag-3'
     (bridging oligo for 5' construction and 5'PCR oligo)

2)   5'-atctatgaaagaattttatatctc-3'
     (bridging oligo for 3' ligation)

3a) 5'-gaattgtggtgttggctccctatagtgagtcgtatta-3'
     (template construction oligo)

3b) 5'-tcagattaaaagtctgatgctccctatagtgagtcgtatta-3'
     (template construction oligo)

4)   5'-tttcatagatnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnncttgattg-3'
     (template construction oligo)

5)   5'-ccggatccgtttcaatagagatataaaattc-3'
     (3' cloning oligo and template construction oligo)

6)   5'-gtttcaatagagatataaaattctttcatag-3'
     (3' primer for PCR)

7)   5'-ccgaagcttctaatacgactcactatagggag-3'
     (5' PCR primer for cloning and for inhibition assay)

8)   5'-agagatataaaattctttcatagnnnntttttcccgnnnnnnnnncggaanncttgattgt-
           cagattaaaagtc-3'
     (random template for SELEX experiment 3) .

9)   5'-gacgttgtaaaacgacggcc-3'
     (3' PCR and RT extension primer for inhibition assay)

TABLE 3   104
starting RNA

5'-gggagcaucagacuuuuaaucugacaaucaag-[-32 n's-]-
                              -aucuaugaaagaauuuuauaucucuauugaaac-3'
isolate

1.1   ucaagAAUUCCGUUUUCAGUCGGGAAAAACUGAACAaucu   (13)

1.2   ucaagCGUAGGUUAUGAAUGGAGGAGGUAGGGUCGUAaucu   (5)

1.3a   ucaagAAUAUCUUCCGAAGCCGAACGGGAAAACCGGCaucu   (1)
1.3b   ------------------G----------------A-----   (1)
1.3c   --------C-------G------------------------   (1)
1.3d   ------------------G--------------------C-- (1)
1.3e   ------------------G--------------------A-- (1)

1.4   ucaagGGCAUCUGGGAGGGUAAGGGUAAGGUUGUCGGaucu   (4)

1.5   ucaagCCCACGGAUGUCGAAGGUGGAGGUUGGGCGGCaucu   (3)

1.6   ucaagAAGAAGAUUACCCAAGCGCAGGGGAGAAGCGCaucu   (2)

1.7   ucaagGAAUCGACCCAAGCCAAAGGGGAUAAUGCGGCaucu   (2)

1.8   ucaagGAUUAACCGACGCCAACGGGAGAAUGGCAGGGaucu   (2)

1.9a   ucaagAGAGUAUCAUC GUGCCGGCGGGAUAUCGGCGaucu   (1)
1.9b   ----------------C-----------------------   (1)

1.10a ucaagUUUGAACAAGCGGAACAUGCACAGCUACACUCaucu   (1)
1.10b -------C--------------------C---------   (1)

1.11   ucaagCUCACGGAUGUCGAAGGUGGAGGUUGGGCGGCAuc   (1)

1.12   ucaagCAUAGACCGCGUAGGGGGAGGUAGGAGCGGCCaucu   (1)

1.13   ucaagCUCUUUCAUAGACCGCGGAGGAGGUUGGGAGaucu   (1)

1.14   ucaagUUCCUAGUAGACUGAGGGUGGGAGUGGUGGAUGucu   (1)

1.15   ucaagCCAAUUACUUAUUUCGCCGACUAACCCCAAGAaucu   (1)

1.16   ucaagGAGGCCAAUUCCAUGUAACAAGGUGCAACUAAUaucu   (1)

1.17   ucaagUGCGUAUGAAGAGUAUUUAGUGCAGGCCACGGaucu   (1)

1.18   ucaagUAAUGACCAGAGGCCCAACUGGUAAACGGGCGGucu   (1)

1.19   ucaagAGACUCCACCUGACGUGUUCAACUAUCUGGCGaucu   (1)

Nucleotides of the fixed regions are shown as lower case letters.

SUBSTITUTE SHEET

TABLE 4   105
Pseudoknot Motif

1.1   ucaagAAUUCCGUUUUCAGUCGGGAAAAACUGAACAaucu   (13)

1.3a   ucaagAAUAUCUUÇCGAAGCCGAACGGGAAAACCGGCaucu   (1)

2.9   ucaagGUUUCCGAAAGAAAUCGGGAAAACUGucu   (1)

2.4a   ucaagUAGAUAUCCGAAGCUCAACGGGAUAAUGAGCaucu   (3)

2.7a   ucaagAUAUGAUCCGUAAGAGGACGGGAUAAACCUCAa-cu   (3)

1.7   ucaagGAAUCGACCCAAGCCAAAGGGGAUAAUGCGGCaucu   (2)

2.11   ucaagUCAUAUUACCGUUACUCCUCGGGAUAAAGGAGaucu   (1)

1.18   ucaagUAAUGACCAGAGGCCCAACUGGUAAACGGGCGGucu   (1)

1.8   ucaagGAUUAACCGACGCCAA-CGGGAGAAUGGCAGGGaucu   (2)

2.1b   ucaagAAUAUAUCCGAACUCGA-CGGGAUAACGAGAAGaGcu   (7)

1.6   ucaagAAGAAGAUUACCCAAGCGCA-GGGGAGAAGCGCaucu   (2)

2.10   ucaagUAAAUGAGUCCGUAGGAGG-CGGGAUAUCUCCAAcu   (1)

1.9b   ucaagAGAGUAUCAUCCGUGCCGG--CGGGAUAUCGGCGaucu   (1)

2.12   ucaagAAUAAUCCGACUCG---CGGGAUAACGAGAAGAGcu   (1)

1.10b   ucaagUUCGAACAAG--CGGAACAUGCACAGCCACACUCaucu   (1)

2.3a   caagUUAAACAUAAUCCGUGAUCUUUCACACGGGAGaucuaugaaaga   (7)

SUBSTITUTE SHEET

TABLE 4 CON'T

2.2b   aaucaagUACCUAGG-UGAUAAAAGGGAGAACACGUGUGa-cu   (1)

2.2b aaucaagUACCUAGGUGAUAAA-AGGGAGAACACGUGUGa-cu   (1)

2.5a ucaagAUAGUAUCCGUUCUUGAUCAUCGGGACAAAUGaucu   (3)

2.6b ucaagUGAAACUUAACCGUUAUCAUAGAUCGGGACAAaucuaugaa   (2)

Nitrocellulose retention motif

1.2   ucaagCGUAGGUUAUGAAUGGAGGAGGUAGGGUCGUAaucuaug   (5)

1.4   aucugacaaucaagGGCAUCUGGGAGGGUAAGGGUAAGGUUGUCGGaucu   (4)

1.5   ucaagCCCACGGAUGUCGAAGGUGGAGGUUGGGCGGCaucu   (3)

1.11   ucaagCUCACGGAUGUCGAAGGUGGAGGUUGGGCGGCAuc   (1)

1.12   ucaagCAUAGACCGCGUAGGGGGAGGUAGGAGCGGCCaucuaug   (1)

1.13   ucaagCUCUUUCAUAGACCGCGGAGGAGGUUGGGAGaucuaugaaaga   (1)

1.14   ucaagUUCCUAGUAGACUGAGGGUGGGAGUGGUGGAUGucuau   (1)

Secondary structures as predicted by the Zuker program are shown with overlined arrows which highlight the inverted repeats indicative of base-pairing.

TABLE 5

107

Clone	Freq.	Stem 1(a)	Loop 1	Stem 2(a)	Loop 2	Stem 1(b)	Loop 3	Stem 2(b)
2.3a	9 *	UCCGUG	A	UCUUUCA	-	CACGGG	AGaucua	ugaaaga
2.9	1	UUCCGA	A	AGA	AA	UCGGGA	AAACUG	ucu
2.7a	4 *	UCCGU	UAA	GAGG	-	ACGGG	AUAAA	CCUC
1.1	13	UUCCG	UU	UUCAGU	-	CGGGA	AAA	ACUGAA
1.3c	5 *	UUCCG	AG	GCCC	AA	CGGG	AAAC	CGGC
1.18	1	ACCAG	AG	GCCG	AA	CUGGU	AUAA	GGGC
2.4a	19 *	UCCG	AA	GCUCA	A	CGGGA	UAAA	UGAGC
2.1b	7 *	UCCG	AA	CUCG	A	CGGG	AUA	CGAG
2.10	1	UCCG	UA	GGAGG	-	CGGG	AUA	UCUCC
1.9b	2 *	UCCG	-	UGCCGG	-	CGGG	AUA	UCGGCG
2.12	1	UUCG	A	CUCG	-	CGGG	ACA	CGAG
1.10b	2 *	UCCG	UUCUU	CA	AG	CGGA	ACAA	UG
2.5a	6 *	CCG	UUA	GAUCAU	-	CGG	GAUAA	AUGauc
2.11	1	CCG	AC	CUCCU	-	CGG	GAGAA	AGGAG
1.8	2	CCG	UUA	GCCA	A	CGG	GACAA	UGGC
2.6b	5 *	CCG	A	UCAUAGU	-	CGG	GAUAAUGC	aucuauga
1.7	2	CCC	AA	GCC	AAA	CGG	GAGAA	GGC
1.6	2	CCC	AA	GCGC	A	GGG	GAGAA	GCGC
2.2b	17 *	CCU	AG	GUG	AUAAA	AGG	GAGAA	CAC

TABLE 6
                                    108
starting RNA

5'-gggagccaacaccacaauuccaaucaag-[-32 n's-]-
                    -aucuaugaaagaauuuuauaucucuauugaaac-3'

isolate

```
2.1a  ucaag  AAUAUA UCCGAACUCGACGGGAUAACGAGAA Gaucu    (3)
2.1b  -----------------------------------------G--   (7)
2.1c  -----CA----------------------------------G--   (1)
2.1d  --------C--------------------------------G--   (1)
2.1e  ---------------------------------G-------G--   (1)
2.1f  -----------------------------------------G--   (1)
2.1g  -----------------------------------------C-C--  (1)
2.1h  --------------A--------------------------G--   (1)
2.1i  -----GU----------------------------------G--   (1)
2.1j  ----------------------------------------A--G--  (1)
2.1k  --------------C--------------------------G--   (1)


2.2a  ucaagUACCUAGGUGAUAAAAGGGAGAACACGUGA acu     (1)
2.2b  -------------------------------------UG---   (13)
2.2c  ------------------------------------A---G---  (2)
2.2d  ----------------------------------------G---  (1)


2.3a  ucaagUUAAACAUAAUCCGUGAUCUUUCACACGGGAGaucu   (7)
2.3b  ------------------------------------------C--  (1)
2.3c  ---------------------------------------A---A--  (1)


2.4a  ucaagUA  GAUAUCCGAAGCUCAACGGGAUAAUGAGCaucu   (3)
2.4b  -----C-AAU---------------------------------   (1)
2.4c  --------------------------------------G---   (1)
2.4d  -----A-------------------------------------   (1)
2.4e  -----U--AU---------------U-.---------------   (1)


2.5a  ucaagAUAGUAUCCGUUCUUGAUCAUCGGGACAAAUGaucu   (3)
2.5b  ------C-----------------------------------   (1)
2.5c  -----U------------------------------------   (1)
2.5d  -------------A----------------------------   (1)


2.6a  ucaagUGAA CUUAACCGUUAUCAUAGAUCGGGACAAa cu   (1)
2.6b  --------A--------------------------u--   (2)
2.6c  --------------------------------u--   (1)
2.6d  --------A-----------------U-----u--   (1)


2.7a  ucaagAUAUG AUCCGUAAGAGGACGGGAUAAACCUCAacu  (3)
2.7b  ----------U-------------------------G---   (1)


2.8  ucaagGGGUAUUGAGAUAUUCCGAUGUCCUAUGCUGUaCcu    (2)


2.9  ucaagGUUUCCGAAAGAAAUCGGGAAAACUGucu    (1)


2.10 ucaagUAAAUGAGUCCGUAGGAGGCGGGAUAUCUCCAAcu     (1)


2.11 ucaagUCAUAUUACCGUUACUCCUCGGGAUAAAGGAGaucu   (1)
```

SUBSTITUTE SHEET

TABLE 6 CON'T   109

2.12	ucaagAAUAAUCCGACUCGCGGGAUAACGAGAAGAGcu	(1)
2.13	ucaagGAUAAGUGCAGGAAUAUCAAUGAGGCAUCCAAaCcu	(1)
2.14	ucaagAUGAGAUAAAGUACCAAUCGAACCUAUCUAAUACGAcu	(1)
2.15	ucaagACCCAUUUAUUGCUACAAUAAUCCUUGACCUCaucu	(1)
2.16	ucaagUAAUACGAUAUACUAAUGAAGCCUAAUCUCGaucu	(1)
2.17	ucaagAACGAUCAUCGAUAUCUCUUCCGAUCCGUUUGucu	(1)
2.18	ucaagACGAUAGAACAAUCAUCUCCUACGACGAUGCAcu	(1)
2.19	ucaagAUAAUCAUGCAGGAUCAUUGAUCUCUUGUGCUaucu	(1)
2.20	ucaagAGUGAAGAUGUAAGUGCUUAUCUCUUGGGACACaucu	(1)
2.21	ucaagCAACAUUCUAUCAAGUAAAGUCACAUGAUaucu	(1)
2.22	ucaagGAUGUAUUACGAUUACUCUAUACUGCCUGCaucu	(1)
2.23	ucaagGGAUGAAAAUAGUUCCUAGUCUCAUUACGACCAcu	(1)
2.24	ucaagUAGUGUGAUAAUGAAUGGGUUUAUCGUAUGUGGCcu	(1)
1.1	ucaagAAUUCCGUUUUCAGUCGGGAAAAACUGAACAaucu	(17)

TABLE 7
starting RNA   110

5'-gggagcaucagacuuuuaaucugacaaucaagNNttccgNNNNNNNNNcgggaaaaNNNN-
                              -cuaugaaagaauuuuauaucucuauugaaac-3'

isolate

3-2   tcaagTAttccgAAGCTCAAcgggaaaaTGAGcta

3-3   tcaagTAttccgAAGCTTGAcgggaaaaTAAGcta

3-6   tcaagGAttccgAAGTTCAAcgggaaaaTGAActa

3-7   tcaagAGttccgAAGGTTAAcgggaaaaTGACcta

3-25   tcaagGAttccgAAGTGTAAcgggaaaaTGCActa

3-50   tcaagTAttccgAGGTGCCAcgggaaaaGGCActa

3-22   tcaagTAttccgAAGGGTAAcgggaaaaTGCCcta

3-8   tcaagTAttccgAAGTACAAcgggaaaaCGTActa

3-13   tcaagGAttccgAAGTGTAAcgggaaaaCGCActa

3-23   tcaagGAttccgAAGCATAAcgggaaaaCATGcta

3-43   tcaggGAttccgAAGTGTAAcgggaaaaAGCActa

3-45   tcaagTAttccgAGGTGTGAcgggaaaaGACActa

3-21   tcaagTAttccgAAGGGTAAcgggaaaaTGACcta

3-9   tcaagTGttccgAGAGGCAAcgggaaaaGAGCcta

3-37   tcaagTAttccgAAGGTGAAcgggaaaaTACActa

3-56   tcaagAGttccgAAAGTCGAcgggaaaaTAGActa

3-58   tcaagATttccgAGAGACAAcgggaaaaGAGTcta

3-39   tcaagATttccgATGTGCAAcgggaaaaTGCActa

3-33   tcaagTAttccgACGTAACAcgggaaaaGTTActa

3-46   tcaagATttccgACGCACAAcgggaaaaTGTGcta

3-52   tcaagTAttccgATGTCTAAcgggaaaaTAGGcta

SUBSTITUTE SHEET

TABLE 7 CON'T   111

3-16   tcaagGGttccgATGCCCAAcgggaaaaGGGCcta

3-34   tcaagAAttccgACGACGAAcgggaaaaACGTcta

3-35   tcaagTAttccgATGTACAAcgggaaaaAGTActa

                                *   *
3-60   tccagCGttccgTAAGTGGAcgggaaaaACCActa

                               *   *
3-27   tcaagAGttccgTAAGGCCAcgggaaaaAGGTcta

3-15   tcaagGAttccgAAAGGTAAcgggaaaaATGCcta

3-18   tcaagAAttccgCTAGCCCAcgggaaaaGGGCcta   (2)

                              *   *
3-31   tcaagAAtt-cgTTAGTGTAcgggaaaaAACActa

3-26   tcaagCGttccgATGGCTAAcgggaaaaATAGcta

                              *   *
3-32   tcaagGAttccgTTTGTGCAcgggaaaaGGCActa

                              *   *
3-54   tcaagAA-tccgTTTGCACAcgggaaaaCGTGcta

3-41   tcaggAA-tccgAGAAGCTAcgggaaaaAGCGActa

3-29   tcaagATttccgAGGTCCGAcgggaaaaTGGTcta

3-20   tcaagTAttccgAAGGAAAAcgggaaaaCCACcta

3-36   tcaagTGttccgAAGGAAAAcgggaaaaCCACcta

3-28   tcaagAAttccgTAAGGGGTcgggaaaaACCctau

3-48   tcaagGAttccgTATGTCCTcgggaaaaAGGActa

3-59   tcaagAGttccgAAAGGTAAcgggaaaaTTACcta

3-12   tcaagTAttccgATAGTCAAcgggaaaaGCGActa

3-30   tcaagTAttccgAGGTGTTAcgggaaaaCACGcta

3-11   tcaagAAttccgTATGTGATcgggaaaaACCActa

TABLE 7 CON'T   112

3-17   tcaagGAttccgATGTACAAcgggaaaaCTGTcta

3-24   tcaagATttccgAAGGATAAcgggaaaaACCGActa

3-51   tcaagAAttccgAAGCGTAAcgggaaaaCATActa

TABLE 8   113

Template Construction:

GGG AGCCA ACACC ACAAU UCCAA UCAAG -(32N)- AUCUA UGAAA GAAUU UUAUA UCUCU AUUGA AAC

Clone	32n Random Region	Downstream Constant Region	ΔGf kcal/mol
	Clones with AUCA loops		
1	CAG AGAUA UCACU UCUGU UCACC AUCA GGGGA	CUAUG AAAGA	-13.0
2	AU AUAAG UAAUG GAUGC GCACC AUCA GGGCG U	AU CUAUG AAAGA	-19.0
3	GGAAU AAGUG CUUUC GUCGA UCACC AUCA GGG	AU CUAUG AAAGA	-17.5
4	UGGAG UAUAA ACCUU UAUGG UCACC AUCA GGG	U CUAUG AAAGA	-13.3
5	UCA GAGAU AGCUC AUAGG ACACC AUCA GGG	AU CUAUG AAAGA	-13.6
6	CUGA GAUAU AUGAC AGAGU CCACC AUCA GGG	AU CUAUG AAAGA	-10.0
7	GGAUU AAUAU GUCUG CAUGA UCACC AUCA GGG	CUAUG AAAGA	-12.6
8	G GGAGA UUCUU AGUAC UCACC AUCA GGGGG CA	CUAUG AAAGA	-12.6
9	A AAUUA UCUUC GGAAU GCACC AUCA GGGCA UGG	CUAUG AAAGA	-10.9
10	G GGAGA UUCUU ACUAC UCACC AUCA GGGGG CA	CUAUG AAAGA	-10.3
11	GGA AUACU·UUCUU UCGAU GCACC AUCA GGGCG	U CUAUG AAAGA	-17.6
12	UCCA AUAGA GUUAG UAGUU GCACC AUCA GGGC	AU CUAUG AAAGA	-11.3
13	GUAU AGAUA GUUCU ACUGA UCACG AUCA CGGG	U CUAUG AAAGA	- 9.7
14	GGAU AUGAU CUUAU GGUAU GCACG AUCA CGGC	AU CUAUG AAAGA	-17.5
15	uUG UCUUU CAUGU AGUAA GCACG AUCA CGGCG	A CUAUGAAAGA	-10.5
16	AGAGC UAGUU CUUGU UUAAG ACACG AUCA CGG	U CUAUG AAAGA	-12.6
17	ACG AGAUU UAUUU AGAUG UCACG AUCA CGGGC	AC CUAUG AAAGA	- 7.8
18	UAAU UGAUA CUUGC AGAGG AUCA CCCUG CUCG	AU CUAUG AAAGA	-10.8
19	AG AGGAC UCAUU AGAGG AUCA CCCUA GUGCG G	U CUAUG AAAGA	-15.0
20	GAGAU AUCAU AAUUC AUUGU UGAGC AUCA GCC	AU CUAUG AAAGA	-12.6
21	UGUAU AGAGC AUCA GCCUA UACAU UGCGU GGC	A CUAUG AAAGA	-12.9
22	GAGA UCAAU AGUAA GGACC AUCA GGCCU GG	CUAUG AAAGA	-14.6
23	UGAG AUAUC UCUAU AGUGU GGAGC AUCA GCCC	AU CUAUG AAAGA	-15.3
24	A UGAGA UAGAU CAUGC UCAGG AUCA CCGGG	CUAUG AAAGA	-11.3
25	AGAG UAUUC UACAU GAUUU GCAUC AUCU GGGCG	UAUG AAAGA	- 9.3
26	GGAUU AAUUC GUCUU UUGAG UGACG AUCA CGC	A CUAUG AAAGA	-13.3
27	A UUGCG UAAUG UUACC AUCA GGAAC ACCGC GU	AU CUAUG AAAGA	-11.4
28	GA GUAAG AUAGC AUCA GCAUC UUGUU CCCGC C	AU CUAUG AAAGA	-14.6
	Clones with ANCA loops		
29	GCGUU AAUUU GGAUU AUAGA UCACC AACA GGG	AC CUAUG AAAGA	- 7.9
30	GAGA UGUUU AGUAC UUCAG CCACC AACA GGGG	U CUAUG AAAGA	-14.2
31	GUCA UACUC UCUUU GUnnU GCACC AACA GGGC	AU CUAUG AAAGA	- 9.4
32	AUAGU AGAGG AACA CCCUA CUAAG UCCCC GCC	A CUAUG AAAGA.	- 9.5
33	CAACA GAGAU GAUAU CAGGA UGAGG ACCA CCC	AU CUAUG A GGA	-11.8
34	AGAUA UAAUU CUCCU CUUGA UGAGC ACCA GCC	AU CUAUG AAAGA	-18.5
35	UAG AGAUA UGAGA UAGUU GCACC ACCA GGGUG	AU CUAUG AAAGA	-16.8
36	AUA UAGGA GAUAU UGUAG UCACG AGCA CGGG	CUAUG AAAGA	-12.5
	Clones with no ANCA loop		
37	UGCGUCACUUAUUGGAACUCUGGGUGGC	A CUAUG AAAGA	-17.7
38	CUGGAGGAGAUUGUGUAAUCGCUUGAACUCC	A CUAUG AAAGA	- 9.7

SUBSTITUTE SHEET

input RNA: 10,726
background: 72

PROTEIN	COUNTS	MOLARITY	% BOUND
Acetylcholinesterase	66	7.3×10^{-6}	0
	88	3.7×10^{-6}	0
	94	3.7×10^{-7}	0
N-acetyl-ß-D-glucosaminidase	86	9.0×10^{-9}	0
	84	4.5×10^{-9}	0
	66	4.5×10^{-10}	0
Actin	70	1.9×10^{-5}	0
	76	9.7×10^{-6}	0
	58	9.7×10^{-7}	0
Alcohol Dehydrogenase	1130	1.4×10^{-5}	10.5
	116	7.0×10^{-6}	1.2
	90	7.0×10^{-7}	0
Aldehyde Dehydrogenase	898	2.1×10^{-5}	8.4
	1078	1.4×10^{-5}	10.1
	846	1.4×10^{-6}	7.9
Angiotensin I, human	284	2.6×10^{-3}	2.6
	74	1.3×10^{-3}	0
	70	1.3×10^{-4}	0

TABLE 9 CON'T 115

Input RNA: 10,726
background: 72

PROTEIN	COUNTS	MOLARITY	% BOUND
Ascorbate Oxidase	2734	2.4×10^{-5}	25.5
	1308	1.2×10^{-5}	12.2
	360	1.2×10^{-6}	3.4
Atrial Natriuretic Factor	4758	1.1×10^{-4}	44.4
	4416	5.5×10^{-5}	41.2
	4176	5.5×10^{-6}	38.9
Bombesin	1578	2.1×10^{-4}	14.7
	650	1.0×10^{-4}	6.1
	116	1.0×10^{-5}	1.1

TABLE 10   116
      sequence
      number   no. of
                                                                isolates

      1a   tcaagc---ATGAAGATACAGCTCCAGATGCTGGACACatct   (1)
      1b   ------G-G---------------T------------------   (1)
      1c   ------GAG---------------T------------------   (9)
      1d   ----CGAG---------------T------------------   (1)
      1e   ------GAG---------------TG-----------------   (1)

      2   tcaagCTTGAGATACAGATTTCTGATTCTGGCTCGCTatct   (5)

      3a   tcaagATGGACTCGGTATCAAACGACCTTGAGACACatct   (4)
      3b   ------------------------------G-------------   (1)

      4a   tcaagATGGCTGGAGATACA-AACTATTTGGCTCGCCatct   (3)
      4b   --------------------A----------------------   (1)
      4c   --------------------G----------------------   (1)

      5   tcaagAAGCCTTGAGATACACTATATAGTGGACCGGCatct   (3)

      6a   tcaagGGTGCATTGAGAAACACGTTTGTGGACTCTGT-atct   (2)
      6b   -----A-----------------------------G--G----   (2)

      7a   tcaagAGCGAAGATACAGAAGACAATACTGGACACGC-atct   (2)
      7b   ------------------------------------A-T----   (1)

      8   tcaagGGGACTCTTTTCAATGATCCTTTAACCAGTCGatct   (2)

      9a   tcaagAAGAGACATTCGAATGATCCCTTAACCGGTTGatct   (1)
      9b   -------------C----------------------------   (1)

      10   tcaagCACGCATGACACAGATAAACTGGACTACGTGCatct   (1)

      11   tcaagACACCTTGAGGTACTCTTAACAGGCTCGGTGatct   (1)

      12   tcaagTTGAGATACCTGAACTTGGGACTCCTTGGTTGatct   (1)

      13   tcaagGGATCTTGAGATACACACGAATGAGTGGACTCGatct   (1)

      14   tcaagATCGAATTGAGAAACACTAACTGGCCTCTTTGatct   (1)

      15   tcaagGCAGCAGATACAGGATATACTGGACACTGCCGatct   (1)

      16   tcaagGGATATAACGAGTGATCCAGGTAACTCTGTTGatct   (1)

      17   tcaagGTGGATTTGAGATACACGGAAGTGGACTCTCCatct   (1)

      18   tcaagAGATAATACAATGATCCTGCTCACTACAGTTGatct   (1)

      19   tcaagGGAGGTATACAGAATGATCCGGTTGCTCGTTGatct   (1)

      20   tcaagAGAAGAATAGTTGAAACAGATCAAACCTGGACatct   (1)

SUBSTITUTE SHEET

TABLE 11   117
MOTIF I

```
aagGGAUCUUGAGAUACACACGA---AUGAGUGGACUCGaucuaugaaa 13    (1)

agGUGGAUUUGAGAUACACGG-------AAGUGGACUCUCCaucuauga 17    (1)

agGGUGCAUUGAGAAACACGU-------UUGUGGACUCUGUaucuauga 6a    (2)

-CGACCUUGAGACACaucu-3'  5'-agAUGGACUCGGUAUCAAA-  3a    (4)

agAUCGAAUUGAGAAACACUA-------ACUGGCCUCUUUGaucuaug 14    (1)

caaucaagUUGAGAUACCUGAA------CUUGGGACUCCUUGGUUGAUc 12    (1)

aagAUGGCUGGAGAUACAAAAC-----UAUUUGG-CUCGCCaucuauga 4a    (3)

aagAAGCCUUGAGAUACACUAU-----AUAGUGGAC-CGGCaucuauga 5    (3)

aaucaagCUUGAGAUACAGAUU-UCUGAUUCUGG-CUCGCUaucuauga 2    (5)

aagACACCUUGAGGUACUCUU-------AACAGG-CUCGGUGaucuaug 11    (1)
```

MOTIF II

```
ucaagGAGAUGAAGAUACAGCUCUA--GAUGCUGGACACaucuauga 1c    (9)

aaucaagAGCGAAGAUACAGAAGACAA--UACUGGACACGCaucuau 7a    (2)

aaucaagGCAGCAGAUACAGGAU-----AUACUGGACACUGCCGAUc 15    (I)

gAGAAGAAUAGUUGAAACAGAUC----AAACCUGGACaucuaugaaa 20    (1)

aucaagCACGCAUGACACAGAUA------AACUGGACUACGUGCAUc 10    (1)
```

SUBSTITUTE SHEET

caaucaagAGAUAAUACAAUGAUCCUGCUCACUACAGUUGaucuaugaaagaauuuuauaucucuau 18 (1)

ucaagAAGAGACAUUCGAAUGAUCCCUU---AACCGGUUGaucuaugaaagaauuuuauaucucuau 9a (1)

ucaagGGGACUCUUUUCAAUGAUCCUUU---AACCAGUCGaucuaugaaagaauuuuauaucucuau 8 (2)

ucaagGGAGGUAUACAGAAUGAUCCGGU---UGCUCGUUGaucuaugaaagaauuuuauaucucuau 19 (1)

aaucaagGGAUAUAACGAGUGAUCCAGGU-AACUCUGUUGaucuaugaaagaauuuuauaucucuau 16 (1)

TABLE 12   119

Clone	Freq.	Stem 1(a)	Loop 1	Stem 2(a)	Loop 2	Stem 2(b)	Loop 3	Stem 1(b)
13	1	GGAUC	UUGAGAUA	CAC	ACGAAUGA	GUG	GACUC	Gaucu
12	1	caaucaag	UUGAGAUA	CC	UGAACUU	GG	GACUC	CUUGGUUG
6a	2	gGGUGCA	UUGAGAAA	CACG	UU	UGUG	GACUC	UGUaucu
3a	4	ACC	UUGAGACA	Caucu	-OPEN-	agAUG	GACUC	GGU
14	1	agAUCGAA	UUGAGAAA	CA	CUAAC	UG	GACUC	UUUGaucu
2	5	agC	UUGAGAUA	CAGA	UUUCUGAU	UCUG	GCCUC	GCU
11	1	CACC	UUGAGGUA	CU	CUUAAC	AG	G-CUC	GGUG
4a	3	agAUGGC	UGGAGAUA	CAAA	CUA	UUUG	G-CUC	GCCaucu
5	3	agA-A-GCC	UUGAGAUA	CACUAU		AUAGUG	G-CUC	GGC-a-ucu
17	1	agGUGGA	UUUGAGAUA	CAC	GGAA	GUG	GAC-C	UCCaucu
1c	9	AGAUG	AAGAUA	CAGC	UCUAGAU	GCUG	GACUC	Caucu
7a	2	agAGCG	AAGAUA	CAG	AAGACAAUA	CUG	GACA	CGC-a-ucu
15	1	gGCAG	CAGAUA	CAG	GAUAUA	CUG	GACA	CUGCC
20	1	AUAG	UUGAAA	CAG	AUCAAAC	CUG	GACA	cuau
10	1	gCACG	CAUGACA	CAG	AUAAA	CUG	GACau	CGUGC
WT	–	GACGCUG	ACGGUA	CA	-OPEN-	UG	GACUA	CAGCGUC
CON	–	GGGACCC	UUGAGAUA	CACGGC	UUCGG	CCGUG	GGCG	GGGUCUC

TABLE 13   120

Clone	Freq.	ΔG_f Consensus Structure	ΔG_f Competing Structure	Information Score Sequence I	Information Score Sequence II	Total Information Score	K_r/K_a [6a]
13	1	-9.9	-9.7	18.25	12.74	30.99	0.576
12	1	-8.9	-10.2	15.79	10.74	26.53	0.298
6a	2	-11.6	-9.9	16.38	12.74	29.12	1.000
6b	2	-	-11.5	16.38	12.74	29.12	0.042
3a	4	-9.5	-8.8	14.79	12.74	27.53	0.405
14	1	-10.1	-10.2	16.38	10.16	26.54	-
2	5	-8.5	-9.1	18.25	3.87	22.12	0.114
11	1	-8.6	-10.9	13.21	6.93	20.14	-
4a	3	-14.2	-15.6	17.40	7.46	24.86	0.564
5	3	-10.4	-10.8	18.25	7.83	26.08	0.567
17	1	-14.8	-12.0	18.25	12.74	30.99	0.247
1c	9	-10.0	-13.2	15.67	11.42	27.09	0.154
7a	2	-6.4	-5.6	13.21	11.42	24.63	-
15	1	-11.1	-10.4	12.62	11.42	24.04	0.455
20	1	-4.8	-6.4	7.39	4.91	12.30	0.191
10	1	-10.2	-8.6	6.60	10.16	16.76	-
8	-	-	-	-	-	-	0.292
9a	-	-	-	-	-	-	0.106
18	-	-	-	-	-	-	0.149
WT	-	-27.0	-26.3	-	-	-	0.180
CON	-	21.1	-19.7	-	-	-	0.429
32n	-	-	-	-	-	-	0.015
Evol. Pop.	-	-	-	-	-	-	0.435

Sequence I = UUGAGAUACA

Sequence II = UGGACUC

Selection. A simple kinetic mechanism for reversible protein-RNA complex formation in a well-mixed solution is written as follows:

$$(1) \qquad [Pf] + [RNAf_i] \quad \overset{k_{+i}}{\underset{k_{-i}}{\rightleftharpoons}} \quad [P{:}RNA_i], \qquad\qquad i = 1, \dots n,$$
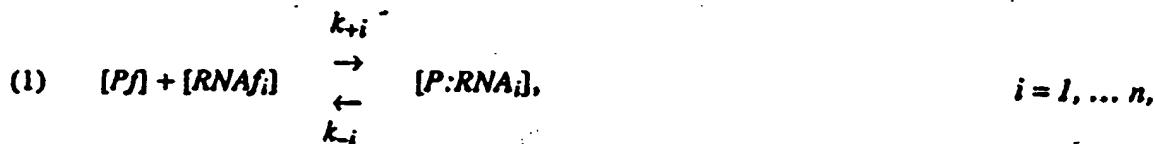
where $[Pf]$ is the free protein concentration, $[RNAf_i]$ is the free RNA species-i concentration, $[P{:}RNA_i]$ is the protein-RNA species-i complex concentration, k_{+i} is the rate constant for association of free protein and free RNA species-i, k_{-i} is the rate constant for dissociation of protein-RNA species-i complexes, and n is the number of RNA sequences with a unique set of rate constants. Alternative mechanisms, including multiple binding sites or cooperativity, could be considered in subsequent treatments with appropriate extensions of this simple scheme.

For any system represented by the above scheme, the fundamental chemical-kinetic or mass-action equations describing the change in concentration of each protein-RNA species-i complex as a function of time are:

$$(2) \qquad \frac{d[P{:}RNA_i]}{dt} = k_{+i} \cdot [Pf] \cdot [RNAf_i] - k_{-i} \cdot [P{:}RNA_i], \qquad\qquad i = 1, \dots n,$$

where $[Pf]$, $[RNAf_i]$, and $[P{:}RNA_i]$ are the concentrations of free protein, free RNA species-i, and protein-RNA species-i complex at time t.

The free protein concentration is the difference between the total protein concentration and the concentration of all protein-RNA complexes $([P] - \Sigma\, [P{:}RNA_i])$; likewise, the free RNA species-i concentration is the difference between the total RNA species-i concentration and the protein-RNA species-i complex concentration $([RNA_i] - [P{:}RNA_i])$:

APPENDIX

SUBSTITUTE SHEET

122

(3) $\quad \dfrac{d[P:RNA_i]}{dt} = k_{+i} \cdot \left([P] - \displaystyle\sum_{k=1}^{n} [P:RNA_k]\right) \cdot ([RNA_i] - [P:RNA_i]) - k_{-i} \cdot [P:RNA_i],$

$$i = 1, \dots n.$$

These dynamic equations can be used for either kinetic or equilibrium analysis. The continuous differential form is valid whenever the mean rate of each process is large relative to the variance in that process, or in other words, Eq. (3) is accurate for description of a pool of RNA with several molecules representing each unique set of rate constants. Whenever there is only one molecule, or just a few molecules of the best-binding RNA present, a statistical description of binding is used to determine the conditions that give a high likelihood of recovering the best-binding RNA. These statistical formulas are derived in a subsequent section on the likelihood of success.

At equilibrium, the change in concentration of each protein-RNA species-i complex equals zero:

(4) $\quad \left([P] - \displaystyle\sum_{k=1}^{n} [P:RNA_k]\right) \cdot ([RNA_i] - [P:RNA_i]) - Kd_i \cdot [P:RNA_i] = 0, \qquad i = 1, \dots n,$

with symbols as defined in Eq. (3), and with Kd_i being the equilibrium dissociation constant for protein-RNA species-i complex ($Kd_i = k_{-i}/k_{+i}$).

When only one RNA species is considered (i.e., $n = 1$), an analytical solution for the equilibrium concentration of protein-RNA complexes is possible by solving the following quadratic equation:

(5) $\quad [P:RNA_1]^2 - ([P] + [RNA_1] + Kd_1) \cdot [P:RNA_1] + [P] \cdot [RNA_1] = 0,$

which has two real roots, one physically realizable:

(6) $\quad [P:RNA_1] = \dfrac{2 \cdot [P] \cdot [RNA_1]}{([P] + [RNA_1] + Kd_1) + \sqrt{([P] + [RNA_1] + Kd_1)^2 - 4 \cdot [P] \cdot [RNA_1]}}$

Of course there are numerous classical approximations for equilibrium or quasi-steady-state concentrations of complexes, like that in the Michaelis-Menten formalism, but none give sufficient accuracy over the range of total RNA and protein concentrations used in SELEX. (For revealing discussions of some pitfalls and limitations of classical approximation see Savageau, 1991; Straus & Goldstein, 1943; Webb, 1963.) Although analytical solution of the quadratic equation for simple reversible association of a single RNA species with a single binding site on the protein is accurate over all RNA and protein concentrations used in SELEX, and although the bound concentrations of two competing species can be calculated by analytical solution of a cubic equation, iterative numerical methods are required to calculate equilibrium concentrations of protein-RNA complexes whenever three or more competing RNA species are considered.

We have developed a computer program to solve for the equilibrium concentration of each protein-RNA species-i complex, $[P:RNA_i]$, given any total protein concentration, $[P]$, any distribution of RNA species-i concentrations, $[RNA_i]$, and any distribution of equilibrium dissociation constants, Kd_i. The Jacobian matrix (e.g., see Leunberger, 1973) for implicit solution of Eq. (4) by Newton's method (e.g., see Leunberger, 1973; Press *et al.*, 1988) is calculated using the following formula:

$$(7) \qquad Fn_i^{eql} = \left([P] - \sum_{k=1}^{n} [P:RNA_k] \right) \cdot ([RNA_i] - [P:RNA_i]) - Kd_i \cdot [P:RNA_i] = 0,$$

$$a_{ij} = \frac{\partial Fn_i^{eql}}{\partial [P:RNA_j]} = -([RNA_i] - [P:RNA_i]) - \delta_{ij} \cdot \left([P] - \sum_{k=1}^{n} [P:RNA_k] + Kd_i \right),$$

$$i = 1, \dots n,$$
$$j = 1, \dots n,$$

where a_{ij} is the element in row-i, column-j of the Jacobian matrix, with $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$.

Often the success of Newton's method depends on a good initial estimate for the solution (e.g., see Leunberger, 1973; Press *et al.*, 1988), in this case, the equilibrium concentration of each protein-RNA species-*i* complex, $[P{:}RNA_i]$. By using the bulk K_d for the total RNA pool, the concentration of protein in all protein-RNA complexes can be estimated:

$$(8) \qquad [P{:}RNA] \approx \frac{2 \cdot [P] \cdot [RNA]}{([P]+[RNA]+<K_d>) + \sqrt{([P]+[RNA]+<K_d>)^2 - 4 \cdot [P] \cdot [RNA]}}$$

where $[P{:}RNA]$ is the concentration of all protein-RNA complexes, $[RNA]$ is the concentration of the total RNA pool, and $<K_d>$ is the bulk equilibrium dissociation constant for the total RNA pool, calculated using the following formula:

$$(9) \qquad <K_d> = [RNA]_{[P]/2} - \frac{[P]}{2} = \frac{1}{\displaystyle\sum_{i=1}^{n} \frac{F_i^0}{K_{d_i} + \frac{[P]}{2}}} - \frac{[P]}{2}$$

where $[RNA]_{[P]/2}$ is the total RNA concentration that binds half the protein and $F_i^0 = [RNA_i]/[RNA]$.

With this estimate for the concentration of protein in complexes, an initial approximation for the concentration of each protein-RNA species-*i* complex can be made using the following formula:

$$(10) \qquad [P{:}RNA_i] \approx \frac{[RNA_i] \cdot ([P] - [P{:}RNA])}{([P] - [P{:}RNA]) + K_{d_i}}, \qquad\qquad i = 1, \dots n.$$

Solutions for the values of $[P{:}RNA_i]$ that satisfy Eq. (4) can be refined to a high level of accuracy by iterative application of Newton's method using Eq. (7). In this implementation, we attain solutions with more than twelve significant digits in less than four or five iterations of Newton's method. This rapid convergence to an accurate solution is due to the initial approximations in Eq. (10) typically giving one or more significant digits at the onset—depending on the range of

SUBSTITUTE SHEET

equilibrium dissociation constants and the abundance of each RNA species. One reason for this level of accuracy is that errors in [P:RNA] tend to cancel in Eq. (10) whenever [P] – [P:RNA] is greater than Kd_i, for example, when [RNA] is less than Kd_i or when Kd_i is less than $<Kd>$. Interestingly, this means that accuracy tends to be higher for any protein-RNA species-i complex with better binding than the bulk RNA pool. Representative examples of the initial accuracy of enrichment calculations—defined as the increase in the fraction of the total RNA pool composed of the best-binding RNA species in each round, and approximated by substituting Eq. (10) into Eq. (20)—

The overall accuracy shown is a reflection of the accuracy of the equilibrium concentrations calculated for every protein-RNA species-i complex using Eq. (10). In a subsequent section, we capitalize on this accuracy to calculate optimum RNA and protein concentrations for maximum enrichment.

Partitioning. Any method of partitioning different species of nucleic acid sequences— including filter binding (Tuerk & Gold, 1990), gel-mobility shifts (Blackwell & Weintraub, 1990), affinity chromatography (Ellington & Szostak, 1990; Green et al., 1990; Oliphant & Struhl, 1987; Oliphant & Struhl, 1988), antibody precipitation, phase partitions, or protection from nucleolytic cleavage (Robertson & Joyce, 1990)—could be used to advantage with SELEX. For example, with filter binding most protein-RNA complexes stick to a nitrocellulose filter while most free RNA molecules wash through (Uhlenbeck et al., 1983; Yarus, 1976; Yarus & Berg, 1967; Yarus & Berg, 1970). The actual fraction of protein-RNA complex that sticks and then can be recovered from the filter is treated in the next section.

Since a fraction of free RNA molecules also sticks to the filter as nonspecific background, the total amount of each RNA species-i collected on the filter is calculated using the following formula, which accounts for both the desired signal from the best-binding RNA molecules in protein-RNA complexes and the noise from free RNA molecules collected as nonspecific background plus competing RNA molecules in protein-RNA complexes:

(11) $RNA_i^{fill} = Vol \cdot \{[P{:}RNA_i] + BG \cdot ([RNA_i] - [P{:}RNA_i])\} \cdot 6.02 \cdot 10^{23}$, $i = 1, \ldots n$,

where RNA_i^{fill} is the number of molecules of RNA species-i collected, Vol is the volume of the reaction mixture passed through the filter, $[P{:}RNA_i]$ is the equilibrium concentration of protein-RNA species-i complex calculated as described in the preceding section, BG is the fraction of free RNA collected as nonspecific background, and $[RNA_i]$ is the total RNA species-i concentration. Any method of partitioning typically gives less than perfect separation of bound and unbound ligands, and hence, requires a measure for the fraction of free ligands collected as background with bound ligands in each round.

As already mentioned, not all protein-RNA complexes in solution may be collected on the filter. Furthermore, RNA in tightly bound complexes may be retained better on the filter than RNA in weakly bound complexes. Whenever this is true, enrichment for RNA molecules that bind tightly would be further enhanced in each round of SELEX. On the other hand, if some molecules could not be eluted from the filter as well as others, their enrichment would be reduced.

Amplification and Renormalization. The amount of each RNA species-i recovered from the filter is calculated using the following formula:

(12) $RNA_i^{pcr} = FR \cdot RNA_i^{fill}$,

$i = 1, \ldots n$,

where FR is the fraction of RNA that can be recovered from the filter, and RNA_i^{fill} is the number of molecules of RNA species-i collected on the filter as calculated with Eq. (11). In this treatment, the value of FR is assumed to be constant and is determined both by the fraction of protein-RNA complex that sticks to the filter and by the fraction of RNA in those complexes that can be recovered and copied by reverse transcriptase to make cDNA for PCR. Assuming that FR is constant for all species is a reasonable starting point, since given sufficient time, when all molecules have the same primer sites for PCR and an excess of primer molecules is used, each species—whether rare or abundant—has virtually the same likelihood of annealing with a primer

molecule. Also, since each RNA molecule is the same length, there is no differential rate of amplification on the basis of size. Of course, if any RNA species has a secondary structure that interferes with primer annealing for cDNA synthesis, or if the primary or secondary structure of the corresponding cDNA slows the rate of DNA polymerase during PCR amplification, enrichment of that species is reduced. We do not incorporate these effects, since there are no good rules to predict what structures actually make a difference. When more is learned about these structures, any significant effects can be added to the mathematical description of SELEX.

The total amount of RNA recovered from the filter is calculated by summing the number of molecules of each species collected to make cDNA copies for PCR amplification:

$$(13) \quad RNA^{pcr} = \sum_{i=1}^{n} RNA_i^{pcr}$$

Any "carrier" or "nonspecific competitor" molecules should be excluded from the total in Eq. (13), since without PCR primer sites these molecules do not amplify. Affinity measurement protocols often include these nonspecific competitor RNA molecules, and if such molecules also are used in SELEX, obviously they should be nonamplifiable. Interestingly, whenever nonspecific competitor molecules interact with the protein at the same site as the best-binding ligand molecules, the main consequence of adding competitor molecules is a reduction in the number of specific sites available for selection. Hence, to determine the protein concentration that binds the desired amount of amplifiable ligand molecules with a high concentration of nonspecific competitor molecules present, corrected binding curves must be generated by including the appropriate concentration of these molecules in each titration. The advantages of using a high concentration of nonspecific, nonamplifiable competitor molecules in each round of SELEX can include a reduction in adsorption of amplifiable ligand molecules to any nonspecific sites on labware, a reduction in binding of amplifiable ligand molecules to any nonspecific sites on the target protein, or a reduction in the fraction of free amplifiable molecules collected as nonspecific background on "false-

SUBSTITUTE SHEET

partitioning" sites—but only when such sites are present in significant numbers and are effectively saturated by the amount of nonspecific competitor molecules used. If these conditions are not met, the effect of adding nonspecific competitor molecules essentially is the same as reducing the amount of protein used.

The amount of each amplifiable RNA species-i recovered after one round, relative to the total in Eq. (13), is calculated as follows:

$$(14) \quad F_i^I = \frac{RNA_i^{pcr}}{RNA^{pcr}}, \qquad\qquad i = 1, \dots n.$$

After PCR amplification of cDNA copies and renormalization of the RNA pool back to its original concentration by *in vitro* transcription (from identical promoter sites on all cDNA molecules) the concentration of each RNA species after one round of SELEX is:

$$(15) \quad [RNA_i] = F_i^I \cdot [RNA], \qquad\qquad i = 1, \dots n,$$

where [RNA] is the total concentration of the RNA pool. For each additional round of SELEX, the concentration of every RNA species can be computed by reiteration of Eqs. (7)-(15), with F_i^I for each RNA species from one round being the starting fraction F_i^0 in the next [see Eq. (9)].

SUBSTITUTE SHEET

<u>CLAIMS</u>

1.   A method for identifying nucleic acid ligands from
a candidate mixture of nucleic acids, said nucleic acid
ligands being a ligand of a given target comprising:
          a)   contacting the candidate mixture with the
target, wherein nucleic acids having an increased
affinity to the target relative to the candidate
mixture may be partitioned from the remainder of the
candidate mixture;
          b)   partitioning the increased affinity
nucleic acids from the remainder of the candidate
mixture; and
          c)   amplifying the increased affinity nucleic
acids to yield a ligand-enriched mixture of nucleic
acids.

2.   The method of claim 1 wherein said candidate
mixture is contacted with said target under conditions
favorable for binding, and nucleic acid-target pairs
contain formed.

3.   The method of claim 1 further comprising the step:
          d)   repeating steps a) through d) using the
ligand enriched mixture of each successive repeat as
many times as required to yield a desired level of
increased ligand enrichment.

4.   The method of claim 3 wherein said level of
increased ligand enrichment is sufficient to establish
a ligand solution to said target.

5.   The method of claim 1 wherein said target is a
protein.

6.   The method of claim 5 wherein said protein is a
nucleic acid binding protein.

130

7.   The method of claim 5 wherein said protein is not known to bind nucleic acids.

8.   The method of claim 1 wherein said amplification step employs polymerase chain reaction (PCR).

9.   The method of claim 1 wherein said partitioning step employs filter binding selections.

10.   The method of claim 1 wherein said candidate mixture is comprised of ribonucleic acids.

11.   The method of claim 9 wherein said contacting step is performed in the presence of an excess of said desired target.

12.   The method of claim 1 wherein said target is supported on a matrix.

13.   The method of claim 12 wherein said matrix-supported target is held in a column.

14.   The method of claim 1 wherein said candidate mixture comprises nucleic acids each comprising a segment of randomized sequence.

15.   The method of claim 1 wherein said candidate mixture is prepared by synthesis from a template containing conserved nucleotides and randomized or biased nucleotides.

16.   The method of claim 1 wherein said candidate mixture comprises nucleic acids each comprising a segment of conserved sequences.

17.   The method of claim 16 wherein said conserved sequence segment of the nucleic acids of said candidate

131

mixture comprises a nucleic acid sequence known to bind
to said target.

18.   The method of claim 2 wherein said target is a
protein and said nucleic acid-target pairs contain
Michael adducts.

19.   A method for preparing a nucleic acid antibody
comprising identifying a ligand solution according to
claim 4, and cloning a nucleic acid derived from said
solution, whereby a purified nucleic acid antibody is
prepared.

20.   A method for selecting a nucleic acid which
affects the function of a target molecule which
comprises the additional step of screening nucleic acid
molecules identified by the method of claim 1 for
increased affinity to said target for their ability to
affect the function of said target molecules.

21.   The method of claim 4 further comprising the steps
of:
        e)   preparing a second candidate mixture
comprising nucleic acids, each nucleic acid comprising
randomized nucleotide segments and conserved nucleotide
segments, and said conserved nucleotide segment
comprising nucleotides sequences derived from said
ligand solution; and
        f)   repeating steps a) through d) using said
second candidate mixture.

22.   The method of claim 4 further comprising the steps
of:
        e)   preparing a second candidate mixture
comprising nucleic acids, each nucleic acid containing
biased nucleotides derived from said ligand solution;
and

132

    f) repeating steps a) through d) using said
second candidate mixture.

  23. The method of claim 1 wherein said target is a
transition-state analog.

  24. A method of making a mixture of nucleic acids
enriched for nucleic acid ligands of a given target
comprising the steps:

    a) preparing a candidate mixture of nucleic
acids comprising a conserved segment and a randomized
segment;

    b) contacting the mixture of nucleic acids
with the target under conditions favorable for binding,
to form nucleic acid-target pairs and unbound nucleic
acids;

    c) partitioning unbound nucleic acids and
nucleic acid-target pairs to separate unbound nucleic
acids from nucleic acid-target pairs; and

    d) amplification of the partitioned nucleic
acids of the nucleic acid-target pairs to yield a
ligand-enriched mixture of nucleic acids.

  25. The method of claim 24 wherein the randomized
segment of said nucleic acids is chemically
synthesized.

  26. The method of claim 24 wherein the randomized
segment of said nucleic acids is synthesized by an
enzyme catalyzed reaction.

  27. The method of claim 24 wherein the randomized
segment of said nucleic acids is made by cleavage of a
naturally-occuring nucleic acid.

  28. The method of claim 24 wherein the randomized
segment of said nucleic acids is a contiguous string of

at least about 15 nucleotides.

29.   The method of claim 24 wherein the randomized segment of said nucleic acids is a contiguous string of at least about 25 nucleotides.

30.   A nucleic acid candidate mixture comprised of nucleic acids, said nucleic acids comprised of a conserved nucleotide segment and a randomized nucleotide segment.

31.   The nucleic acid candidate mixture of claim 30 wherein said randomized nucleotide segment is a contiguous string of at least about 15 nucleotides.

32.   The nucleic acid candidate mixture of claim 30 wherein said randomized nucleotide segment is a contiguous string of at least about 25 nucleotides.

33.   The nucleic acid candidate mixture of claim 30 wherein said mixture contains at least about 10^9 nucleic acids.

34.   The nucleic acid candidate mixture of claim 30 wherein said mixture contains at least about 10^{14} nucleic acids.

35.   The nucleic acid candidate mixture of claim 30 wherein said randomized nucleotide segment is flanked by conserved nucleotide segments.

36.   The nucleic acid candidate mixture of claim 35 wherein said flanking conserved nucleotide segments are selected to aid in the amplification of selected nucleic acids.

37.   The nucleic acid candidate mixture of claim 30

wherein said conserved and randomized segments are arranged to enhance the percentage of nucleic acids in said candidate mixture having a given tertiary structure.

38. The nucleic acid candidate mixture of claim 37 enhanced in nucleic acids configured as hairpin loops.

39. The nucleic acid candidate mixture of claim 37 enhanced in nucleic acids configured as pseudoknots.

40. A synthetic nucleic acid antibody having a specific binding affinity for a target, such target being a three dimensional chemical structure.

41. The nucleic acid antibody of claim 40 comprising a nucleic acid ligand selected from the group consisting of single-stranded RNA, double-stranded RNA, single-stranded DNA, double-stranded DNA and chemical modifications thereof.

42. The nucleic acid antibody of claim 40, said target being a structure other than a polynucleotide which binds to said nucleic acid antibody through a mechanism which predominantly depends on Watson/Crick base pairing or triple helix binding; provided, however, that when the nucleic acid antibody is double stranded DNA, the target is not a naturally occuring protein whose physiological function depends on specific binding to double stranded DNA.

43. The nucleic acid antibody of claim 40 wherein the target is a protein.

44. The nucleic acid antibody of claim 40 wherein the antibody is comprised of single stranded RNA.

SUBSTITUTE SHEET

45. The nucleic acid antibody of claim 40 comprised of a plurality of target-specific ligands.

46. The nucleic acid antibody of claim 45 comprised of more than one identical ligands.

47. The nucleic acid antibody of claim 45 comprised of at least two distinct target-specific ligands.

48. The nucleic acid antibody of claim 40 further comprised of a non-nucleic acid element.

49. The nucleic acid antibody of claim 48 wherein said non-nucleic acid element is capable of directing said antibody to a selected location in the body of a patient.

50. The nucleic acid antibody of claim 48 wherein said non-nucleic acid element is a target-specific ligand.

51. The nucleic acid antibody of claim 40 wherein said target is a protein and said antibody is capable of forming a Michael adduct with said target.

52. The nucleic acid antibody of claim 40 comprised of a nucleic acid ligand obtained by the process of claim 1.

53. The nucleic acid antibody of claim 52 wherein said target is a protein.

54. The nucleic acid antibody of claim 52 wherein said nucleic acid ligand is selected from the group consisting of single-stranded RNA, double-stranded RNA, single-stranded DNA, double-stranded DNA and modifications thereof.

136

55.   The nucleic acid antibody of claim 43 wherein said protein is a nucleic acid polymerase.

56.   The nucleic acid antibody of claim 55 wherein said protein is a DNA polymerase.

57.   The nucleic acid antibody of claim 56 wherein said protein is pg 43.

58.   The nucleic acid of claim 57 comprising the RNA sequence

      5'-NNNGAGCCUAGCAACCUGGGCUAGGAAU-3'

or the corresponding DNA sequence thereof or complementary sequences thereof.

59.   The nucleic acid antibody of claim 55 wherein said protein is a reverse transcriptase.

60.   The nucleic acid antibody of claim 59 wherein said reverse transcriptase is HIV-1 reverse transcriptase.

61.   The nucleic acid antibody of claim 60 wherein said antibody is configured as a pseudoknot.

62.   The nucleic acid antibody of claim 61 comprising the RNA sequence 5'-UUCCG-3'.

63.   The nucleic acid antibody of claim 61 comprising the RNA sequence
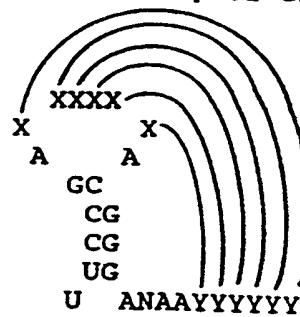
or the corresponding DNA sequences thereof or complementary sequence thereof.

SUBSTITUTE SHEET

137

64. The nucleic acid antibody of claim 61 comprising the RNA sequence 5'-CGGGA-3'.

5

65. The nucleic acid antibody of claim 61 wherein loop 1 is two nucleotides long.

66. The nucleic acid antibody of claim 61 wherein Stem 2 is 5 or 6 base pairs.

10

67. The nucleic acid antibody of claim 61 wherein if Loop 2 exists, the nucleotides are As.

68. The nucleic acid antibody of claim 61 wherein Loop 3 comprises at least 3 nucleotides, said nucleotides enriched in A.

15

69. The nucleic acid antibody of claim 61 wherein Loop 1, Stem 2(a) and Loop 2 consist of eight nucleotides.

20

70. The nucleic acid antibody of claim 43 wherein said protein is a bacteriophage coat protein.

71. The nucleic acid antibody of claim 70 wherein said protein is bacteriophage R17 coat protein.

25

72. The nucleic acid antibody of claim 71 comprising the RNA sequence

30
```
            U   C
          A       A
            CG
            CG
          A
            CG
            NN'
35          NN'
```

or the corresponding DNA sequence thereof or complementary sequences thereof.

40

SUBSTITUTE SHEET

73.   The nucleic acid antibody of claim 43 wherein said protein is a serine protease.

74.   The nucleic acid antibody of claim 73 wherein said protein is tPA.

75.   The nucleic acid antibody of claim 43 wherein said protein is a mammalian receptor.

76.   The nucleic acid antibody of claim 75 wherein said protein is human growth factor.

77.   The nucleic acid antibody of claim 43 wherein said protein is a mammalian hormone or factor.

78.   The nucleic acid antibody of claim 77 wherein said protein is insulin.

79.   The nucleic acid antibody of claim 43 wherein said protein is not known to bind nucleic acids.

80.   The nucleic acid antibody of claim 77 wherein said protein is Nerve Growth Factor.

81.   The nucleic acid antibody of claim 80 comprising a RNA sequence selected from the group consisting of:
        5'-CUCA-3';
        5'-GAGCGCAAGACGAAUAG-3';
        5'-UACA-3'; and
    5'-ACAUCGAUGACCGGAAUGCCGCACACAGAG-3'
    or the corresponding DNA sequence thereof or complementary sequences thereof.

82.   The nucleic acid antibody of claim 56 wherein said protein is HSV-1 DNA polymerase.

83.   The nucleic acid antibody of claim 82 comprising

the RNA sequence:

      5'-UAAGGAGGCCAC-3'

or the corresponding DNA sequence thereof or complementary sequences thereof.

84. The nucleic acid antibody of claim 43 wherein said protein is a ribosomal protein.

85. The nucleic acid antibody of claim 84 wherein said protein is E. coli ribosomal protein S1.

86. The nucleic aic antibody of claim 78 comprising the RNA sequence:

```
                        R N G
                   Y           G
                (C/G)            A
                (U/A)    A
                    C  C
                    N-N'
                    N-N'
                    A-U
                    A-U
                    G-c
   5'--NNYR(G/C)(A/U)GACAC-GNN--3'
```
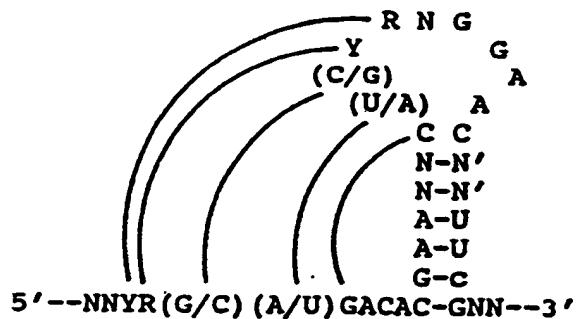
or the corresponding DNA sequence or complementary sequences thereof.

87. The nucleic acid antibody of claim 43 wherein said protein is a viral rev protein.

88. The nucleic acid antibody of claim 87 wherein said protein is HIV-1 rev protein.

89. The nucleic acid antibody of claim 88 configured as an assymetric bulge.

90. The nucleic acid antibody of claim 89 comprising a RNA sequences selected from the group consisting of:

140

```
                              G-A-G-A
                        U            N
                      U      /////     A
        5'--X-X-X-X-X        /////        C-A-X--3'
        3'--Y-Y-Y-Y-Y-C-U-C-A-G-G-U-Y--5'        ;
```

5

10

```
                              G  A
                         N          N
                       N              A
        5'--X-X-X-X-G              C-A-G--3'
        3'--Y-Y-Y-Y-C-A-C-A-G-G-U-C--5'       ; and
```

15

```
         C-C      U-U-G-A-U-C-U-A
        A    G-G           /////    U-G-A-A--3'
        A    C-C    /////    G-C-U-U--5'
         U-U      C-U-A-G-U-A-A
```
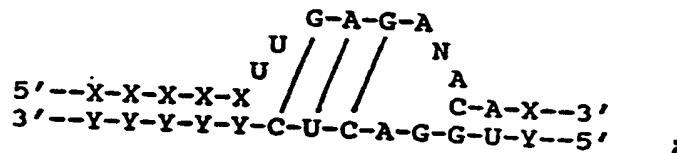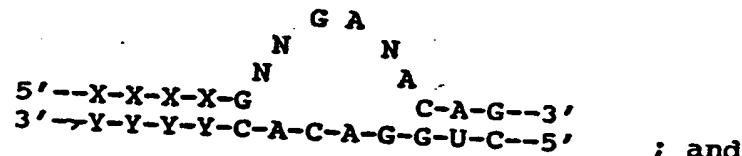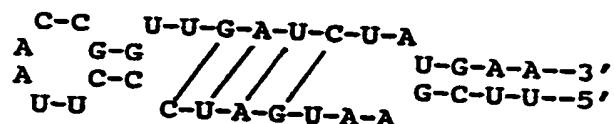
or the corresponding DNA sequences thereof or
complementary sequences thereof.

20

25

30

35
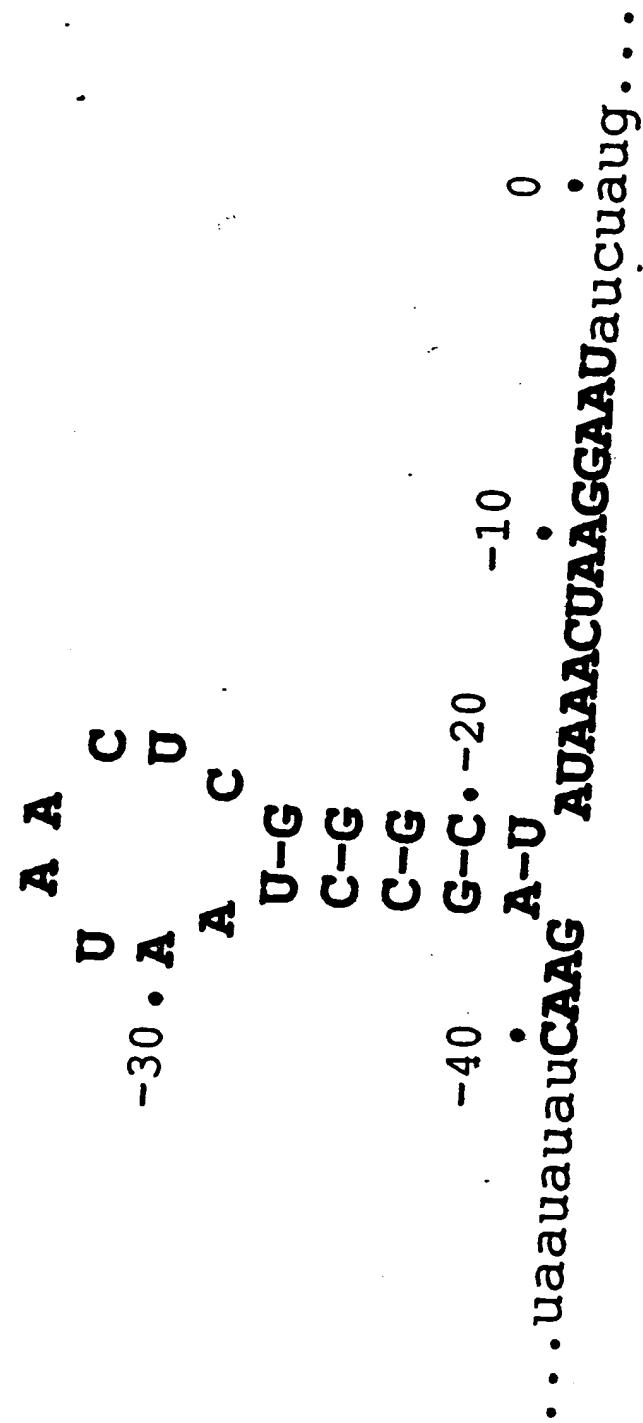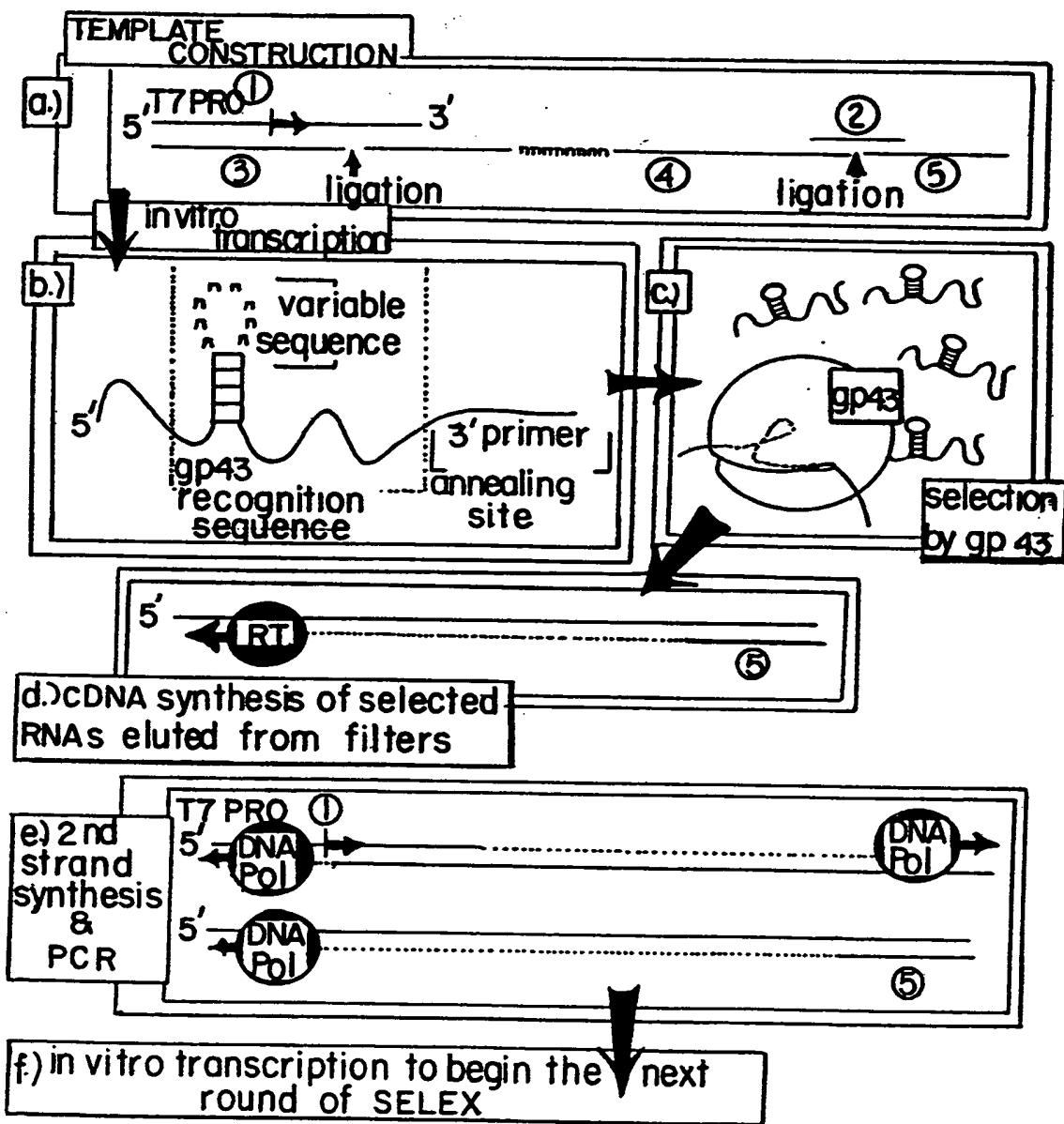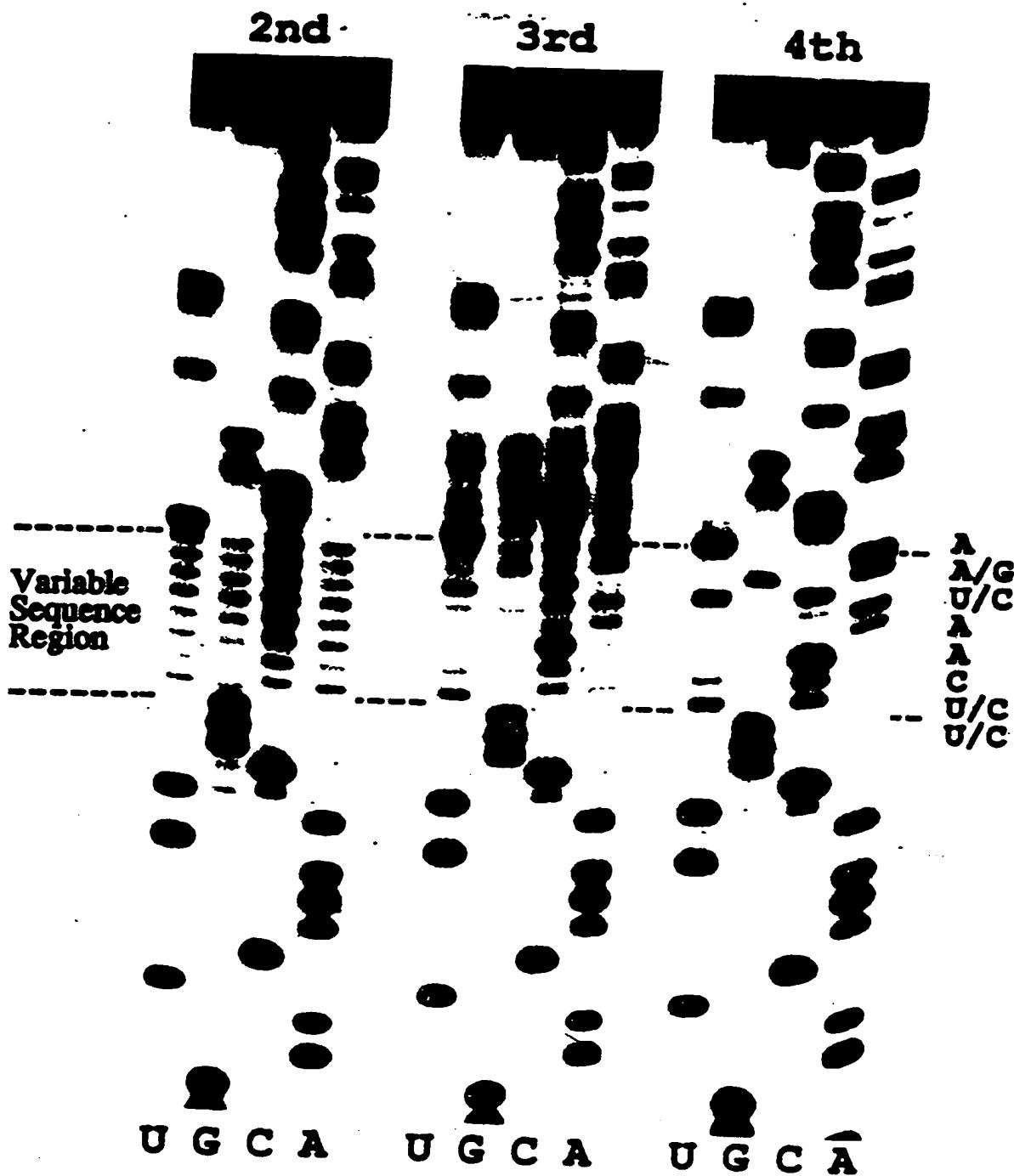
FIG.1

FIG.2

FIG.3

SELECTION CYCLE:

2nd   3rd   4th

Variable
Sequence
Region

A
A/G
U/C
A
A
C
U/C
U/C

U G C A   U G C A   U G C A

FIG.4

FIG.5

6/31

FIG.6

7/34

```
        N  N                          A  A
    N        N  N              U/C         C
  N        N        N        A/G        U/C
     N  U  G  N                  A  U/C
        C  G                        U
        C  G                        C  G
        G  C          SELEX         C  G
     ...GA UA...        ⊥  ✗→       G  C
                                 ...GA UA...
    Kd = 3.2 × 10⁻⁷
```

Kd = 3.2×10^{-7}

```
                    ┌──→ G (1/20)                      ┌──→ G (1/20)
                    │    Kd = 2.7 × 10⁻⁸                │   Kd =
                 ┌─┐                                 ┌─┐    1.7 × 10⁻⁸
                 │A│  A                              │A│  A
    U      └─┘     C                  C    └─┘     C
  A               C                  G          C
  A       ┌─┐     U                  A          U
    A     │C│→ U (1/20)              U          G
    U  G  └─┘    Kd = 1.2 × 10⁻⁸     C          G
    C  G                             C          G
    C  G                             G          C
    G  C                          ..GA UA..
 ..GA UA..
```

$Kd = 2.7 \times 10^{-8}$

$Kd = 1.2 \times 10^{-8}$

$Kd = 1.7 \times 10^{-8}$

Wild Type   Major Variant

(9/20)   (8/20)

⊢———— Kd = 4.8×10^{-9} ————⊣
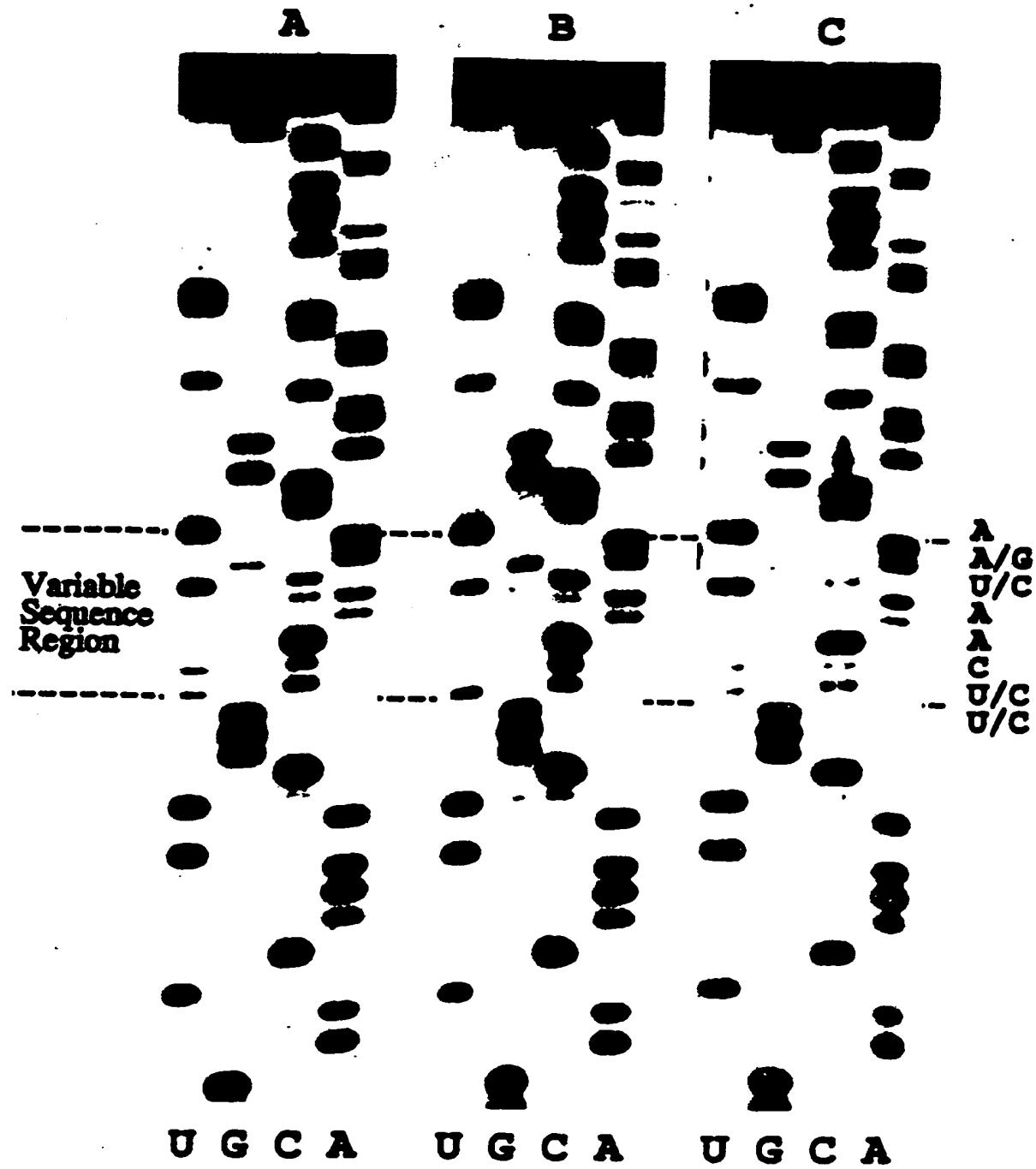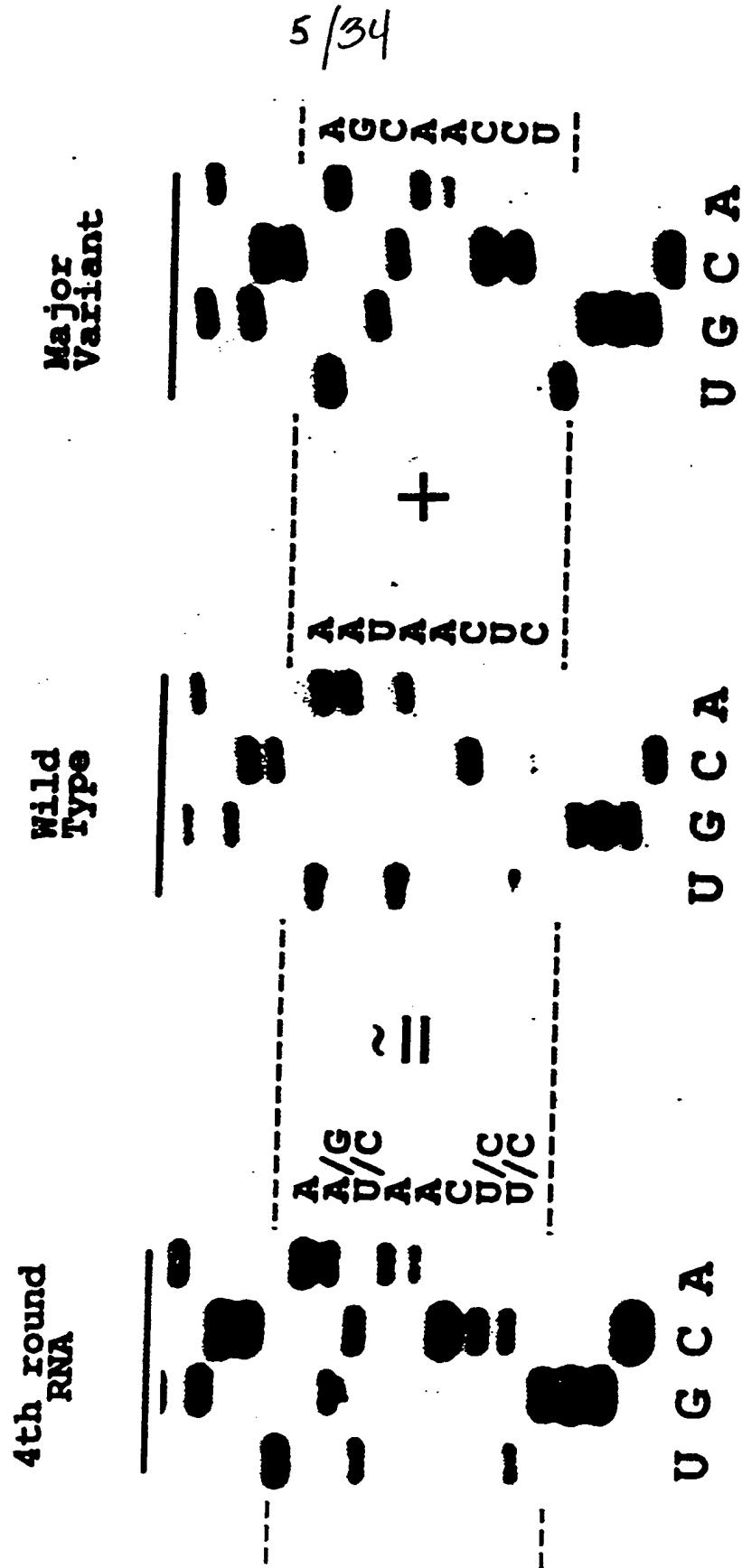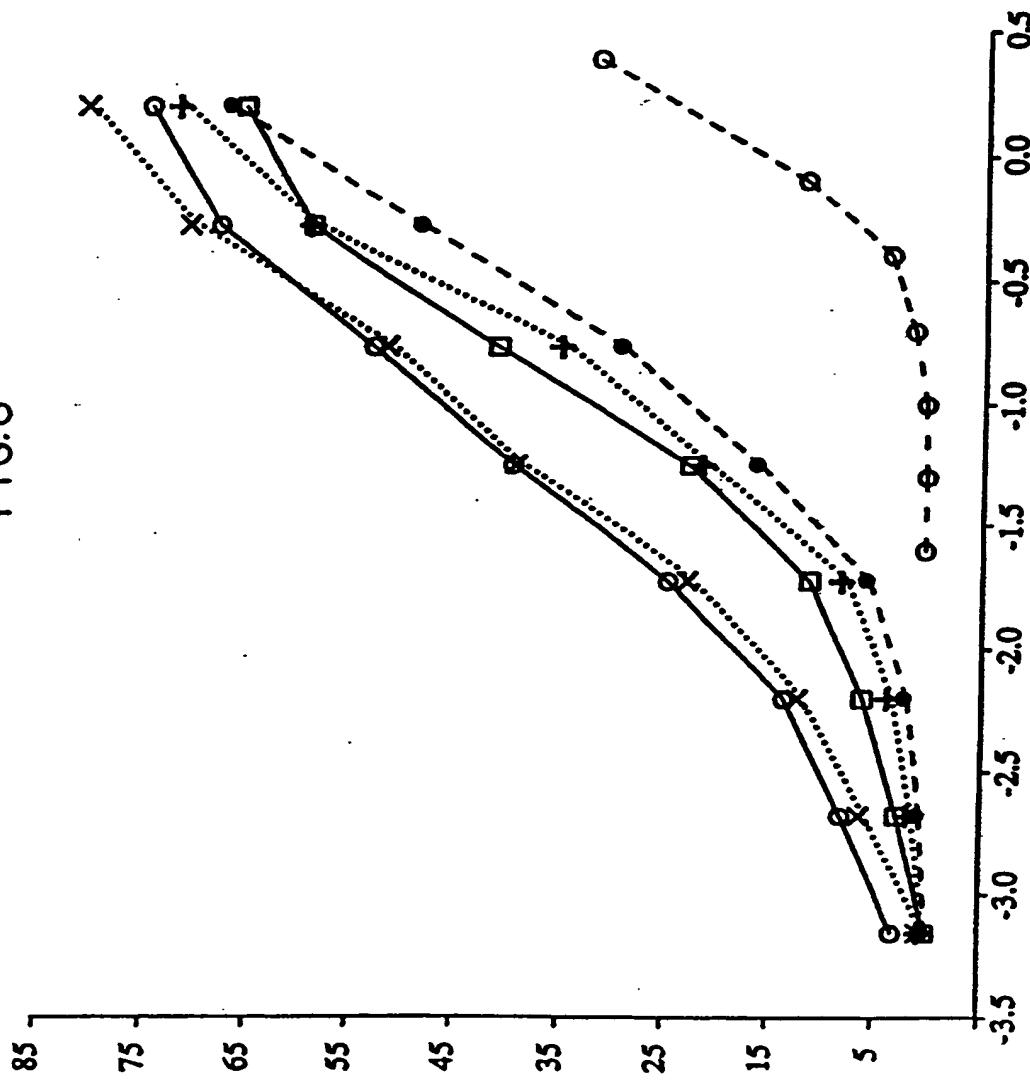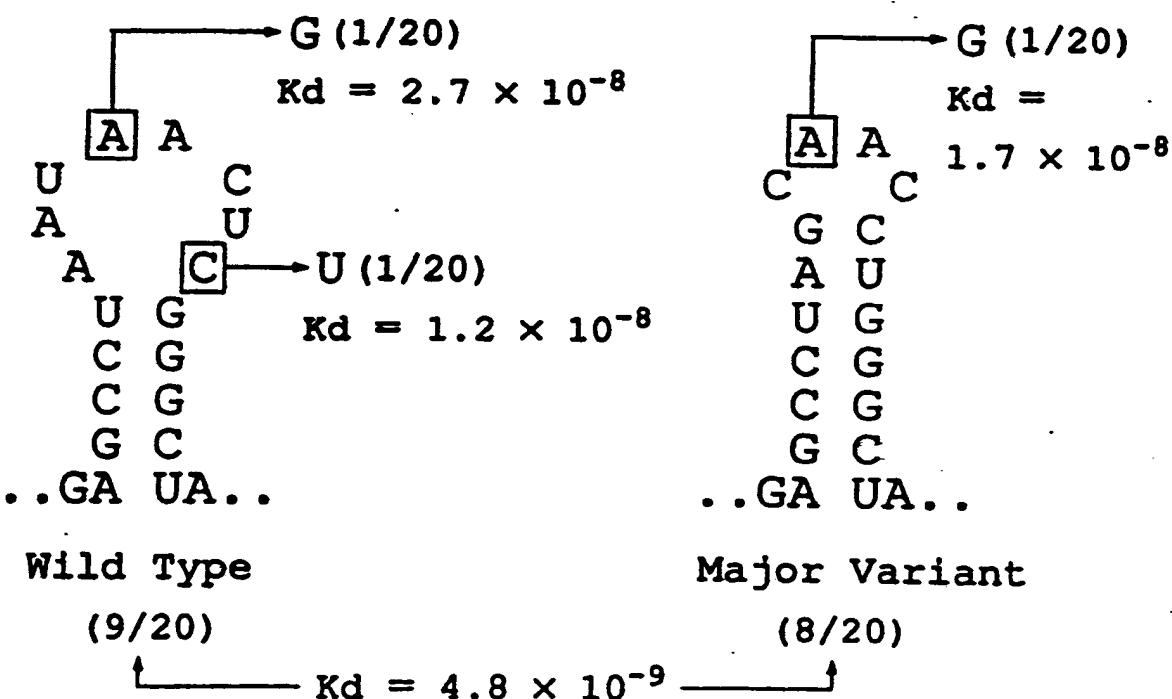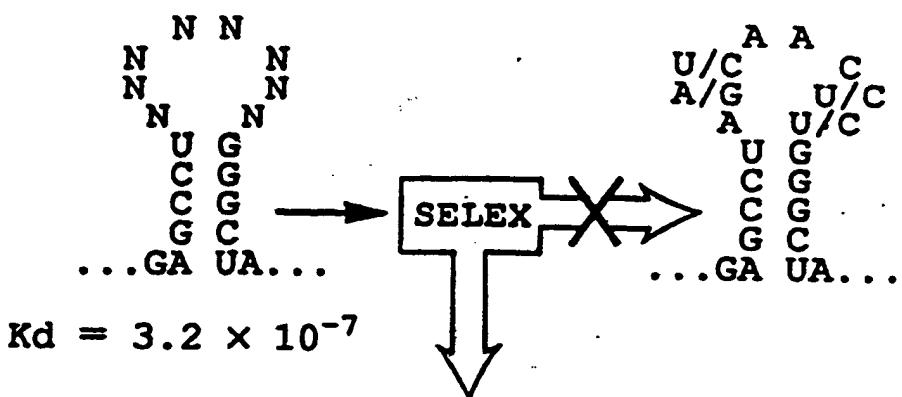
FIG.7

VARIABLE TEMPLATE SYNTHESIS USING TERMINAL TRANSFERASE

5' PRIMER
(OR PRIMARY LIGAND SEQUENCE)　　　　　　3' PRIMER

5' ——————————— 3'　　3' ——————— 5'

TAILING WITH TERMINAL TRANSFERASE
USING RANDOM NUCLEOTIDES.　　　　　　　　　　dNTPs

5' ————————— · · · · · · · · · · · · · · · · · · (T DT) →

HOMOPOLYMER TAILING OF LENGTHENED 5'
PRIMER & 3' PRIMER.　　　　　　　　　　　　　dCTP

5' ——————— · · · · · · · · · · · · CCCCC(T DT) →

GGGGGG ——————
3' PRIMER

ANNEALLING & FILLIN

5' ———————— · · · · · · · · · · · · · · · CCCCCC · · · · · (DNA Pol) →
← (DNA Pol) · · · · · · · · · · · · · · · · · GGGGGG ————

5' ———————— · · · · · · · · · · · · · · · · C C C C C C · · · · · · ——
· G G G G G ————

FIG . 8

"WALKING" BY EXTENDING THE PRIMARY LIGAND.
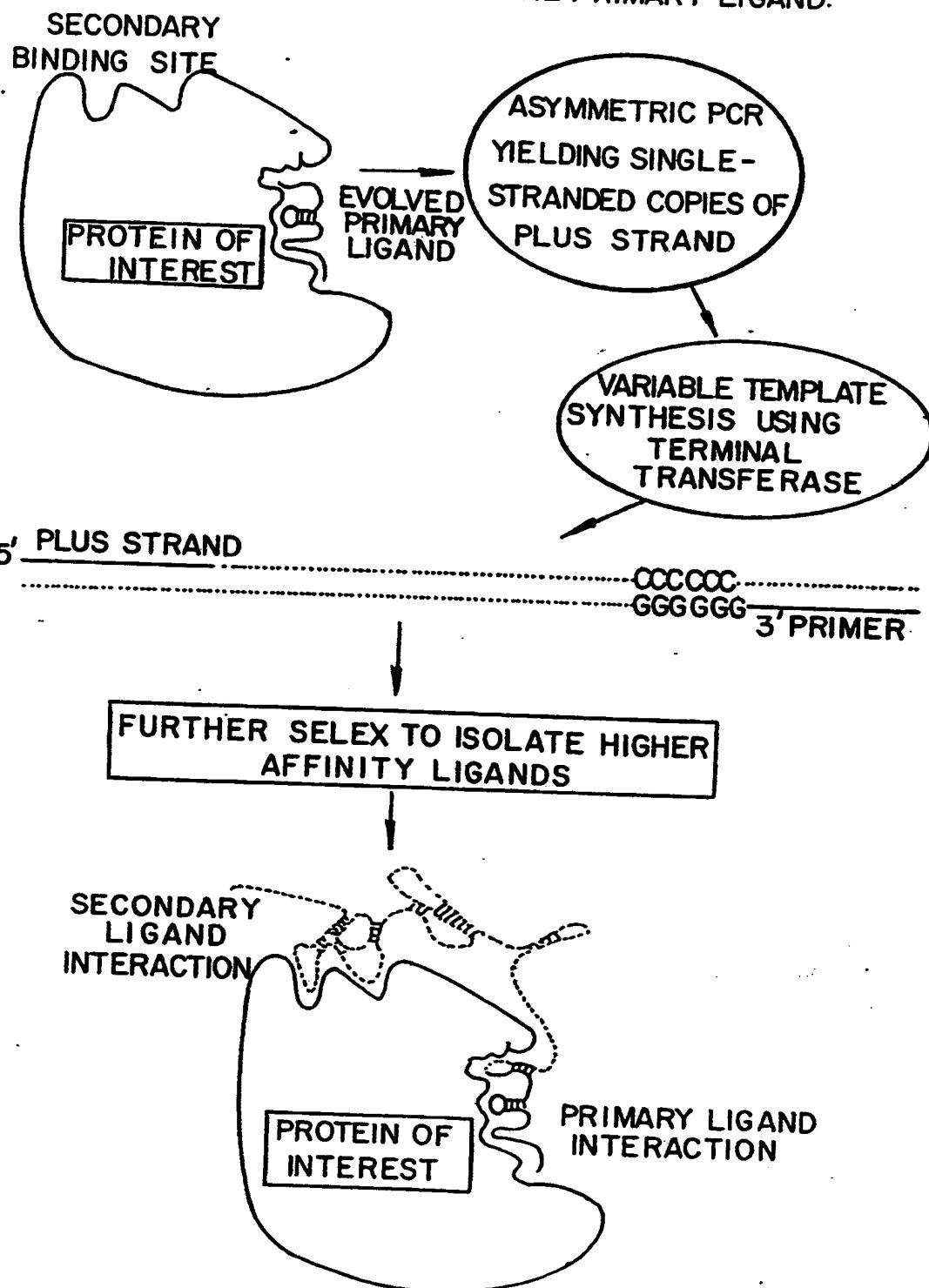
FIG.9

ANCHORING OF BRIDGING OLIGONUCLEOTIDE &
SECONDARY LIGAND EVOLUTION.

GUIDE OLIGONUCLEOTIDE
ANCHORED AT TARGET SITE

PLUS STRAND TO BE USED IN SELEX

SELEX LEADS TO
ISOLATION OF LIGANDS
THAT INTERACT WITH
SECONDARY BINDING
SITES
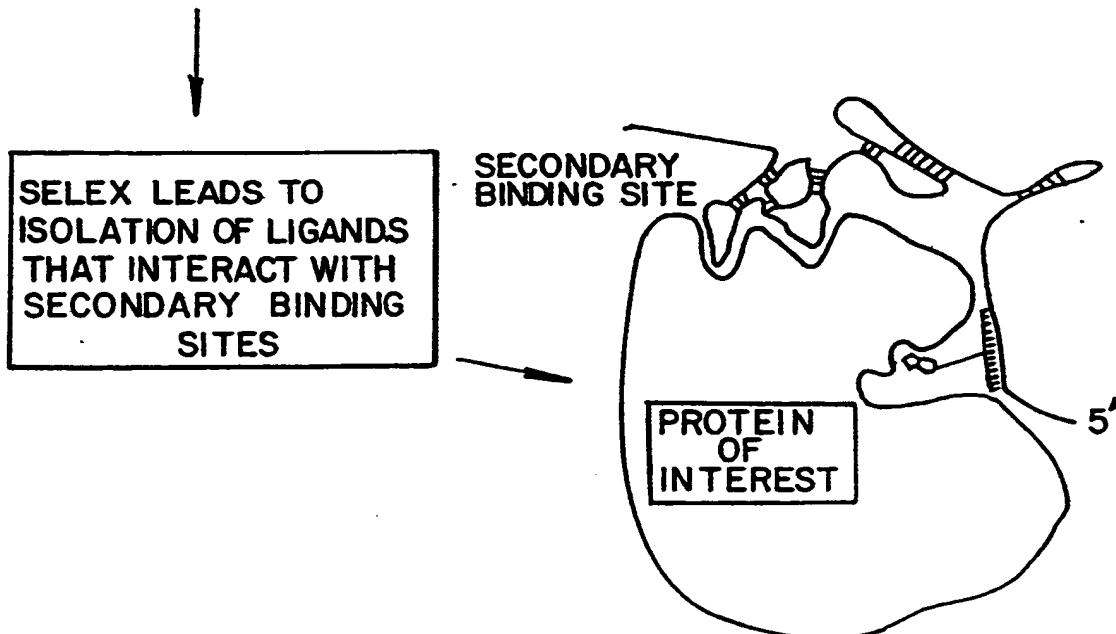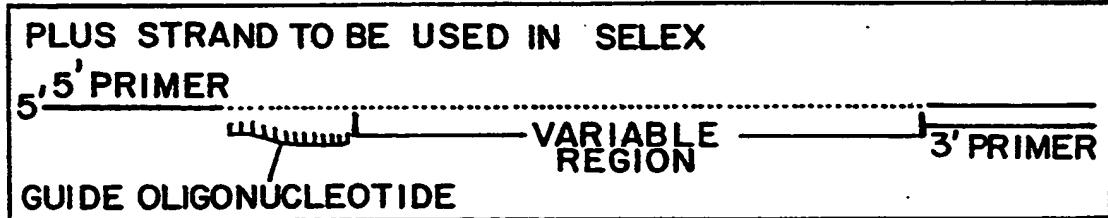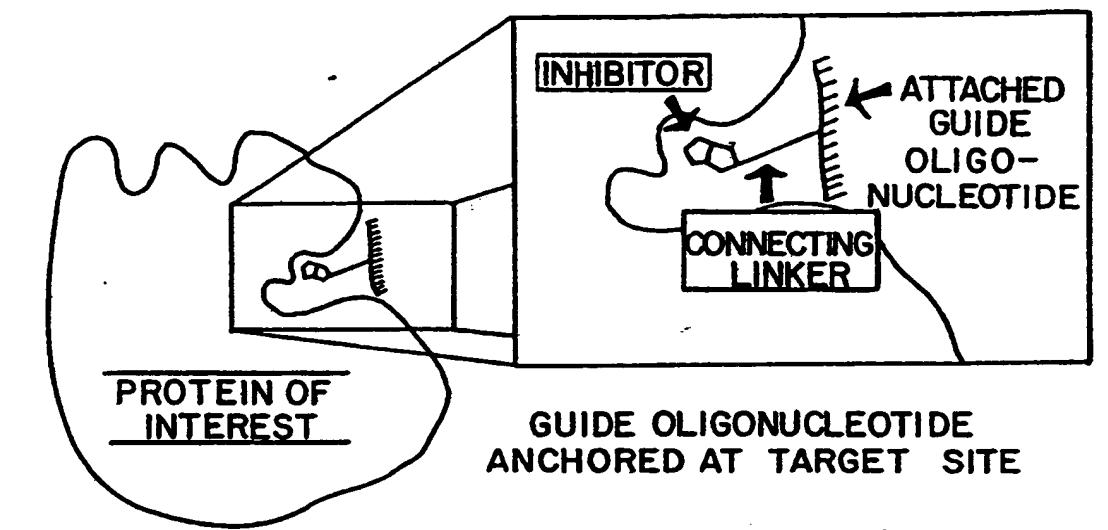
FIG.10

SECONDARY LIGAND-DIRECTED PRIMARY LIGAND EVOLUTION.
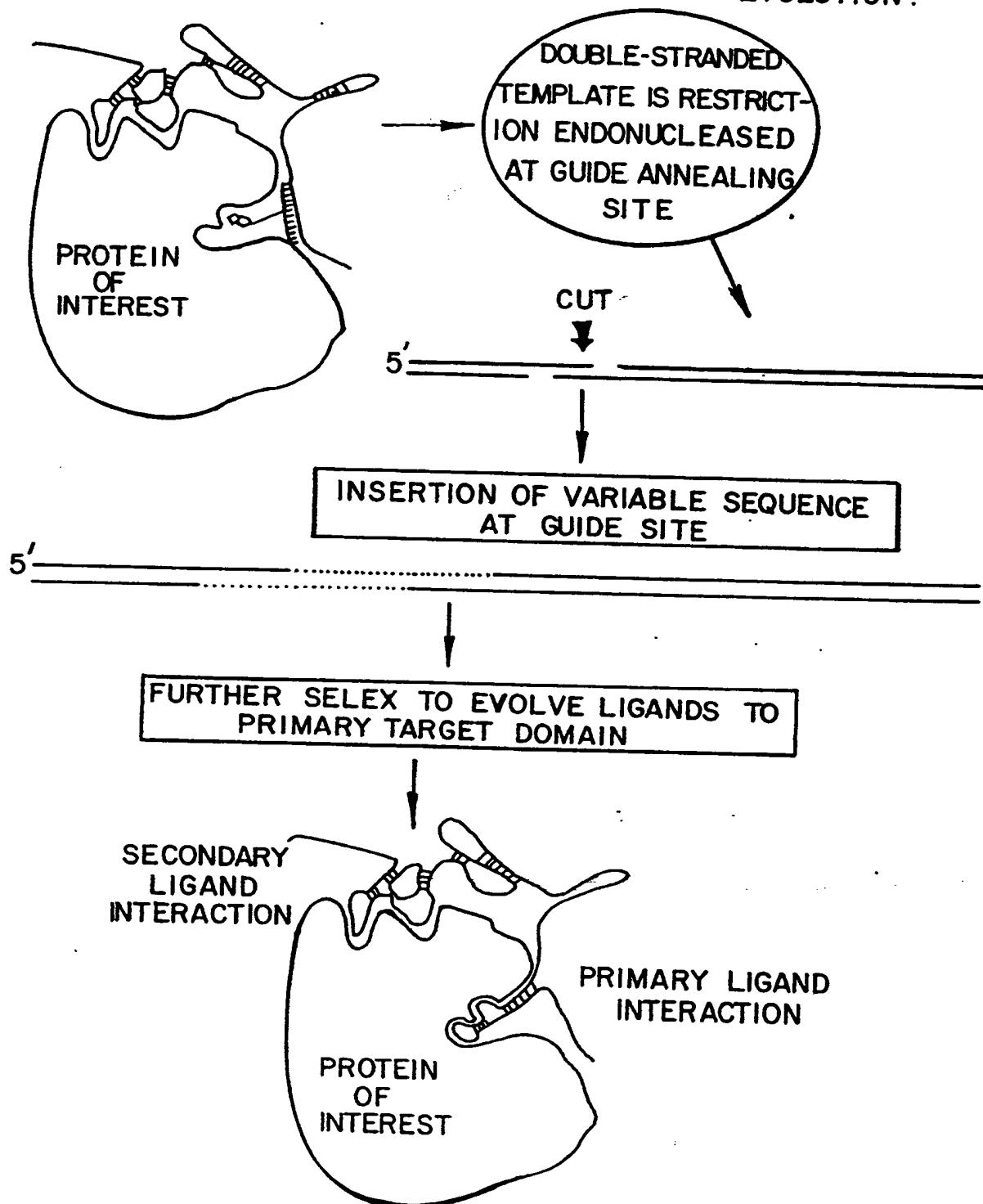
FIG.11

T7 promoter

5'-taatacgactcactataggggagcatcagacttttaatctgacaatcaag-3'

3'-attatgctgagtgatatccctcgtagtctgaaaattagact-5'

ligation

3'-gttagttc....

FIG.12B

FIG.12B

in vitro transcript

5'-ggggagcaucagacuuuuaaucugacaaucaag[32n]aucuaugaaagaauu

→

FIG.I2 B

FIG.I2A

FIG.12A
+

5'-atctatgaaagatttatatctc-3'
3'-cttaaatatagagataactttgcctaggcc-5'

              ←——— ligation

32n..tagatactt-5'

FIG.12B

uuauaucucuauugaaacggauccgg-3'
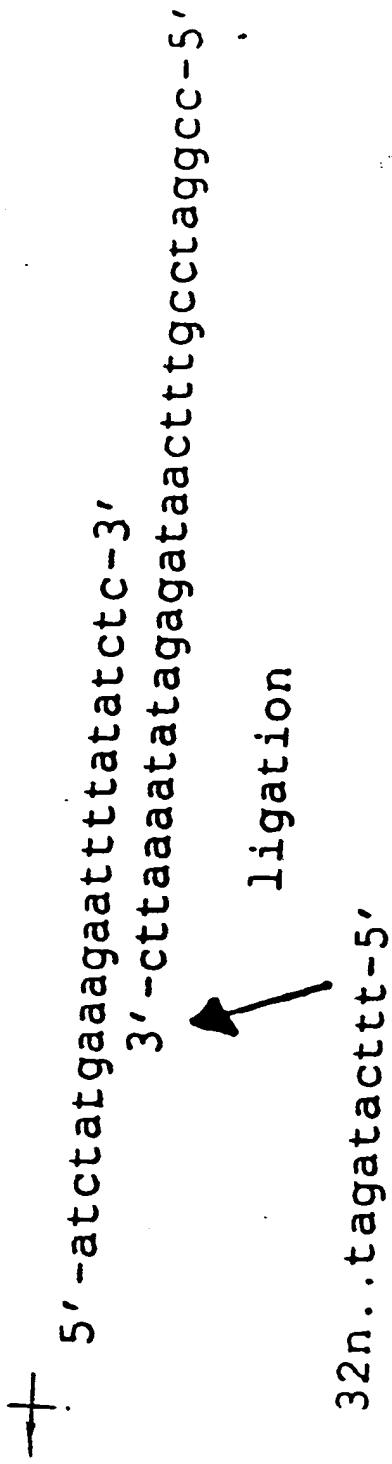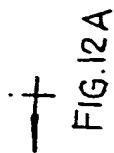
FIG.12A
+
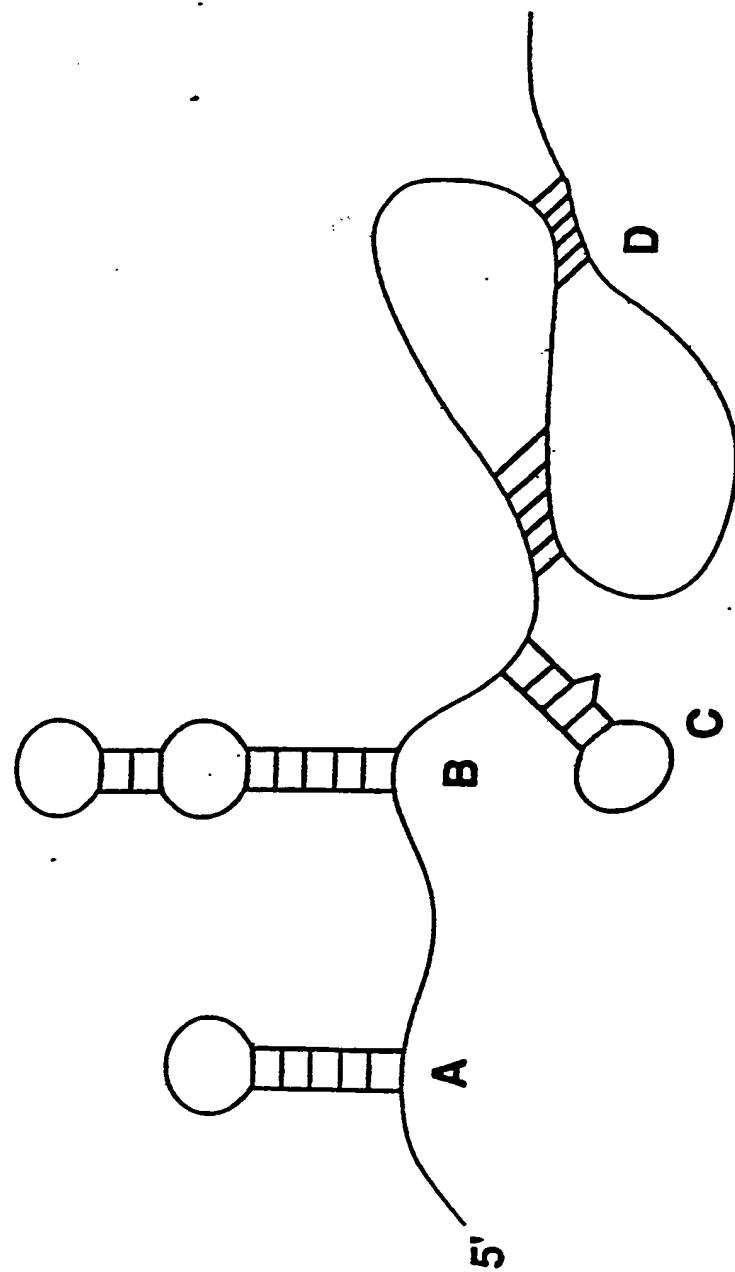
FIG.13

15./34

FIG.14
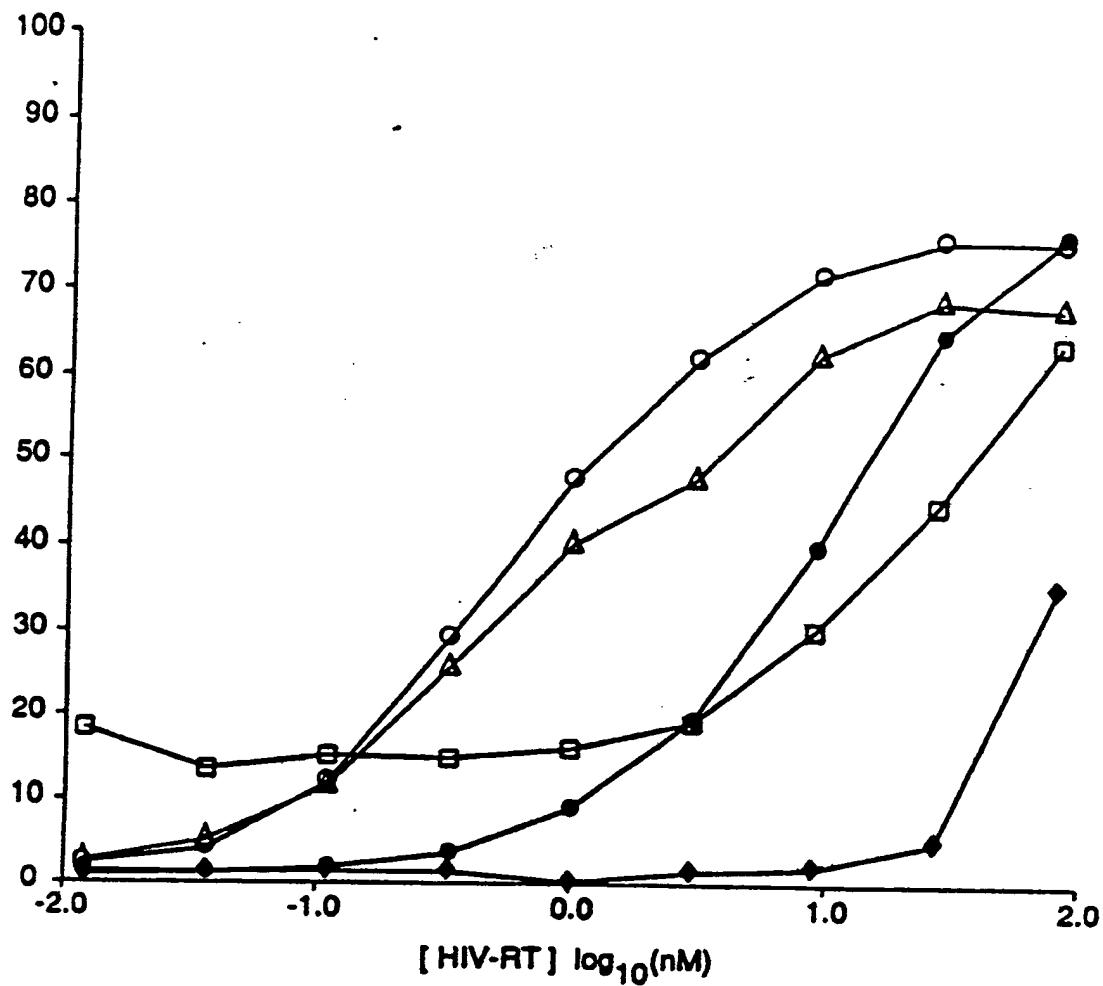
△ 1.1 ucaagAAUUCCGUUUUCAGUCGGGAAAAACUGAACA aucu (13)

O 1.3 ucaagAAUAUCUUCCGAAGCCGAACGGGAAAACCGG Caucu (1)

● 1.3 ----------------G-----------------A----- (1)

☐ 1.4 ucaagGGCAUCUGGGAGGGUAAGGGUAAGGUUGUCGGaucu (4)

16/34

FIG. 15
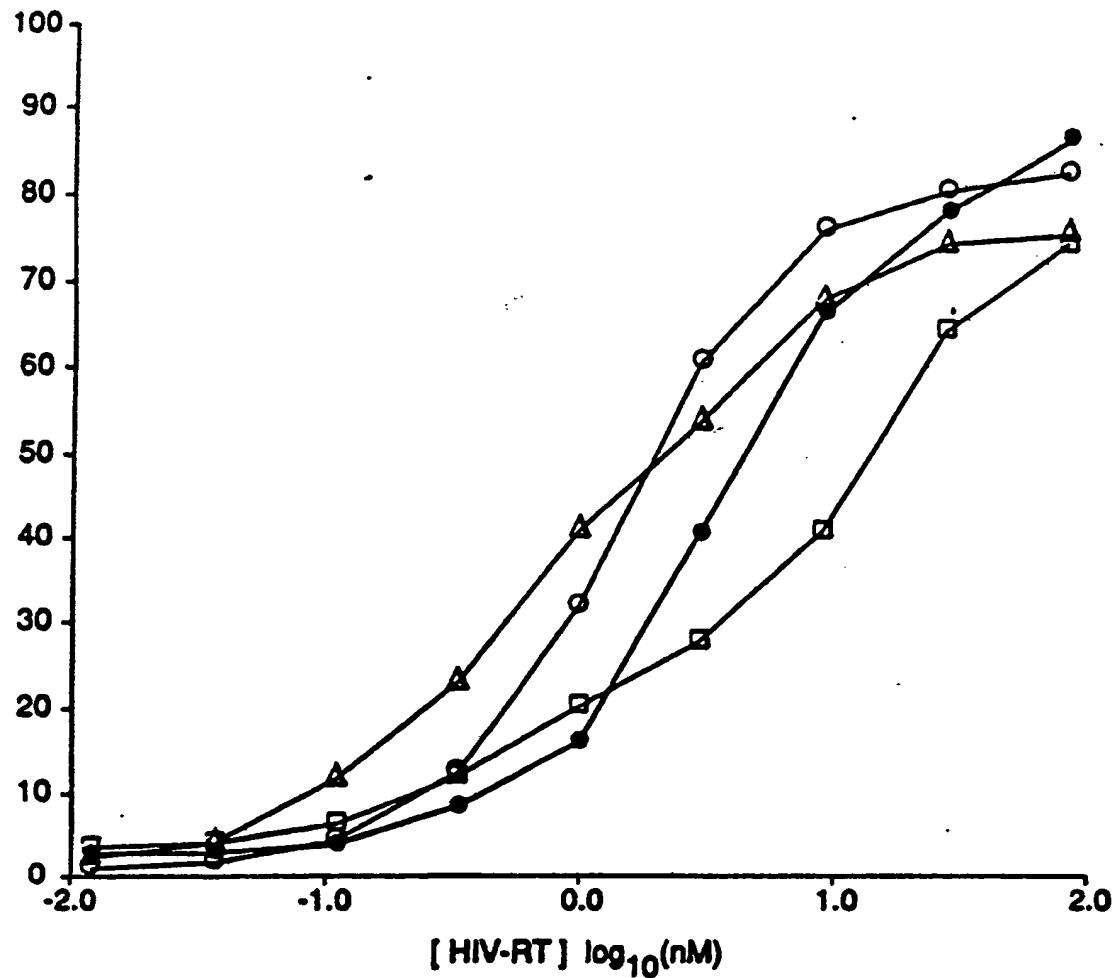
ISOLATE

○ 2.1a ucaag--AAUAUA-UCCGAACUCGACGGGAUAACGAGAA-Gaucu (3)

□ 2.2b ucaaguACCUAGGUGAUAAAAGGGAGAACACGUGUGa-cu   (13)

● 2.5b ucaagACAGUAUCCGUUCUUGAUCAUCGGGACAAAUGaucu   (3)

△ 1.1   ucaagAAUUCCGUUUUCAGUCGGGAAAAACUGAACAAUcu   (13)
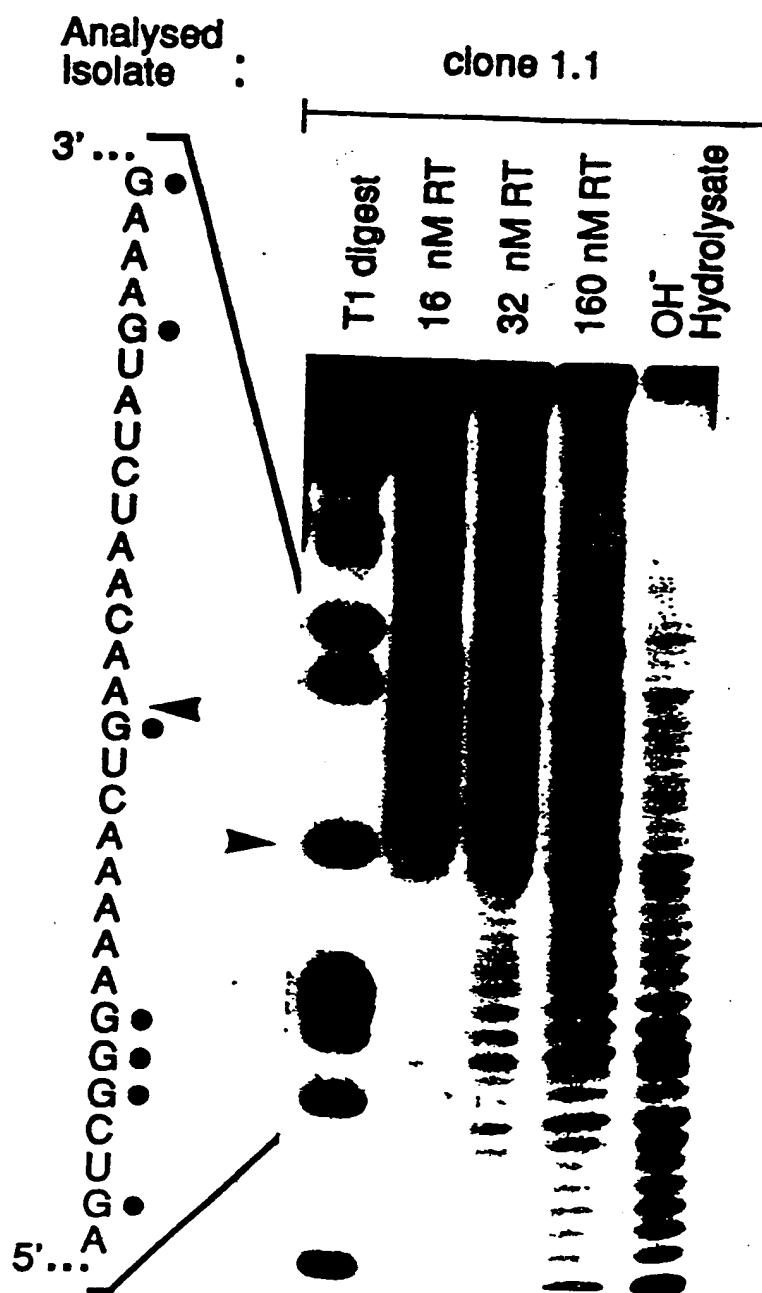
FIG. 16 A

FIG. 16 B

FIG.16 C

FIG.17

FIG.18

FIG.19

U — C

A — A

C — G

C — G

A — C — G

C — G

N — N'

N — N'

Pu — Py'

N — N'

N=N'

N=N'

5' ... 3'
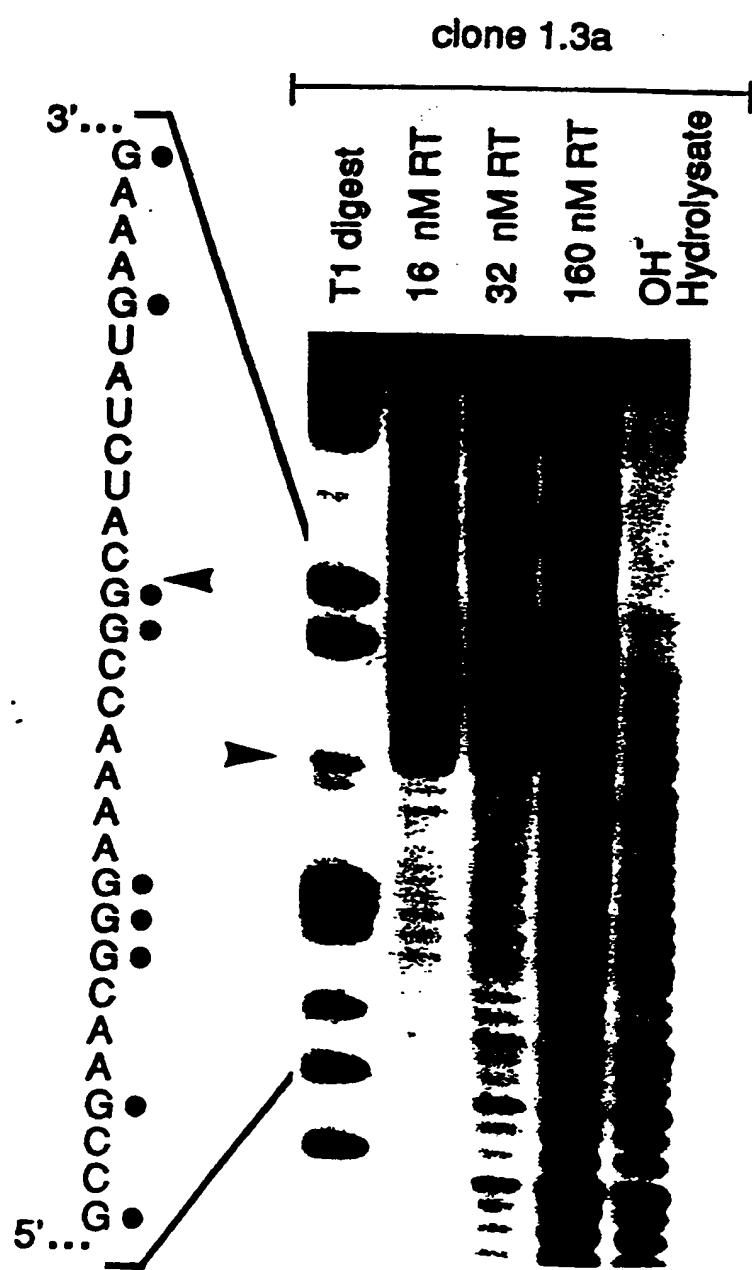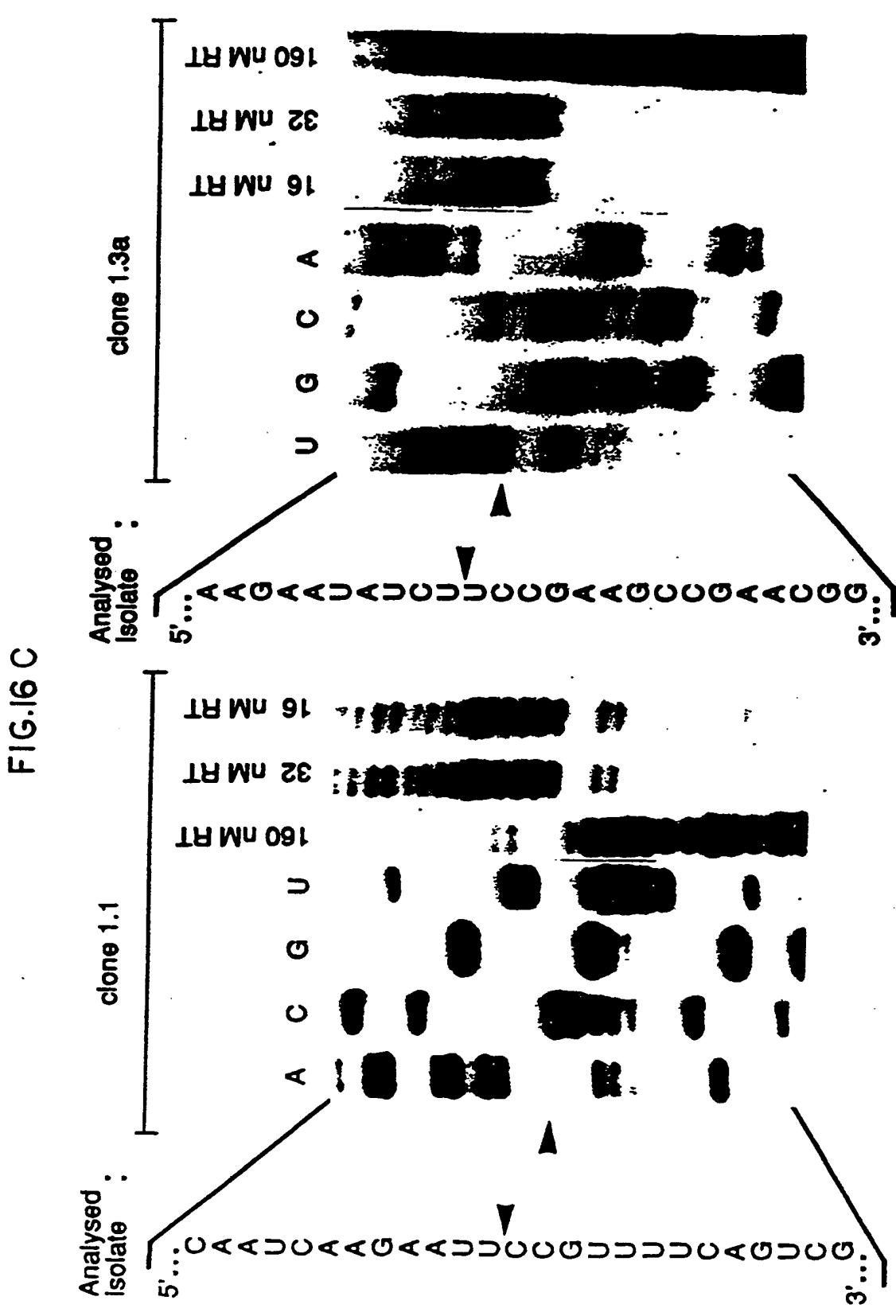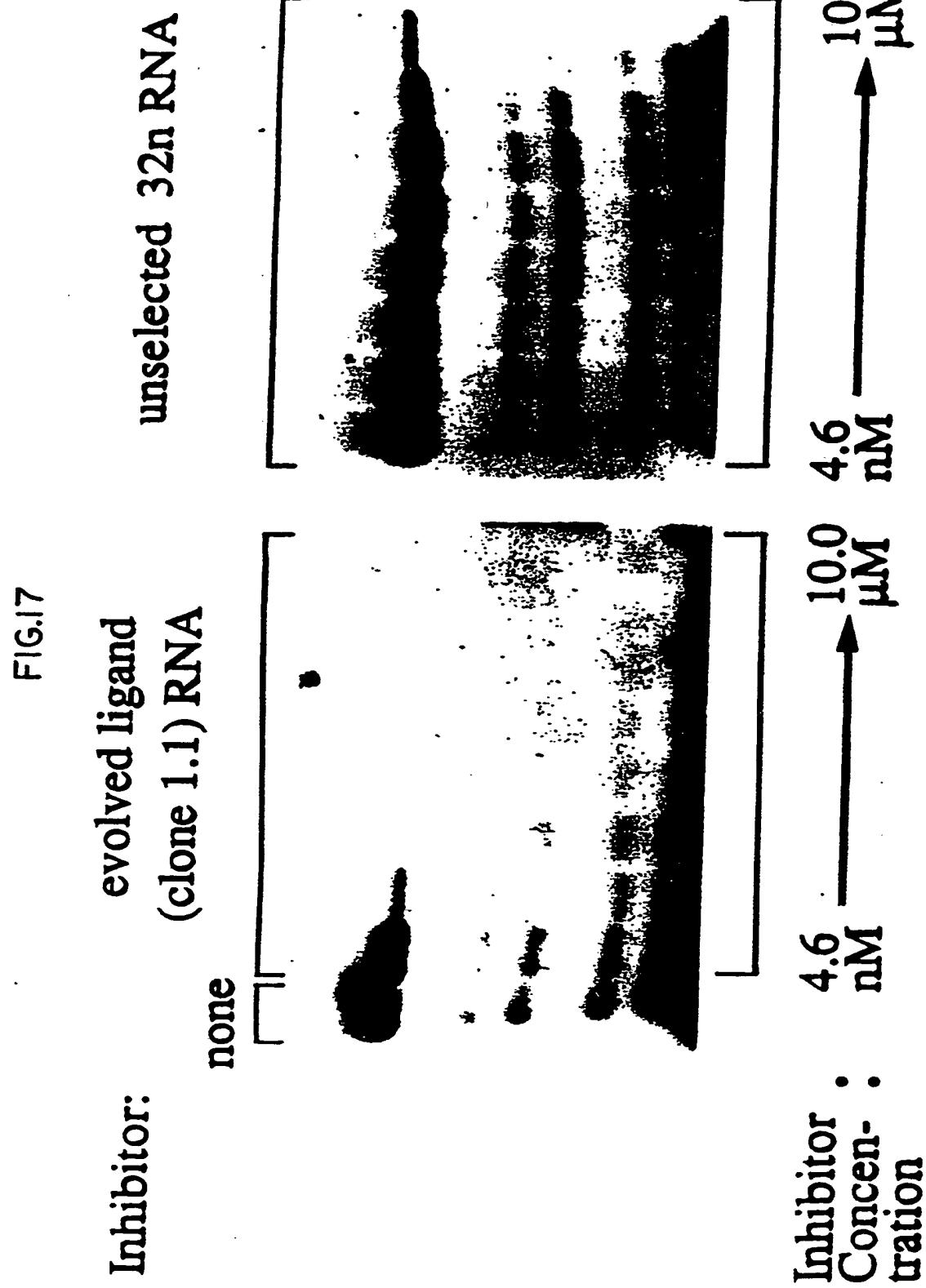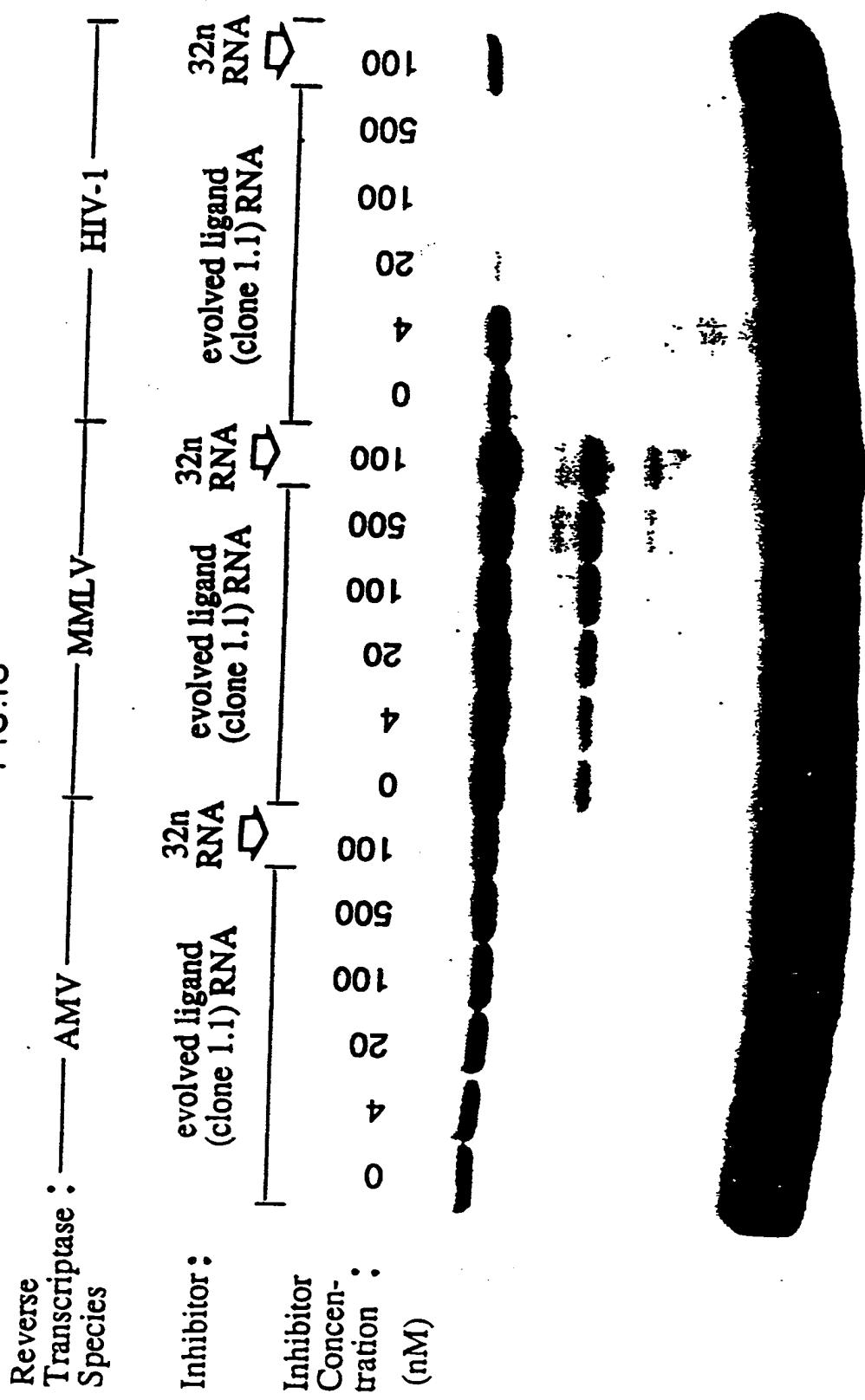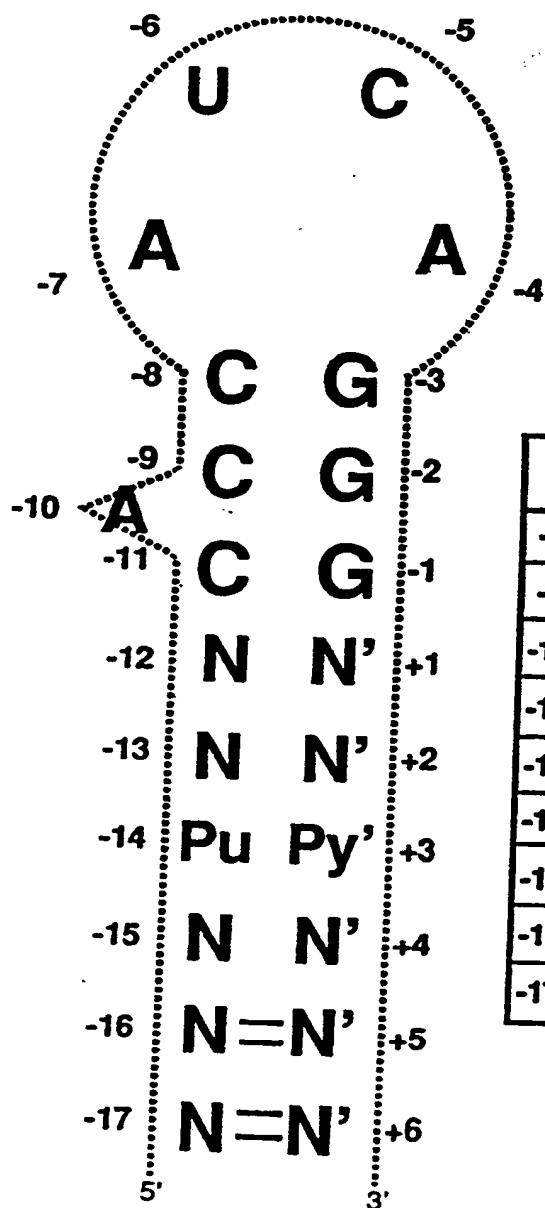
	A	C	G	U
-4	36	0	0	0
-5	0	36	0	0
-6	4	3	1	28
-7	36	0	0	0
-10	36	0	0	0

	AU	CG	UA	GC	UG	GU	Bulge	END
-8/-3	0	24	0	12	0	0	0	0
-9/-2	0	25	0	10	1	0	36	0
-11/-1	0	24	2	10	0	0	0	1
-12/+1	8	1	8	10	7	1	3	3
-13/+2	6	5	8	9	3	1	3	4
-14/+3	9	0	4	10	2	3	6	8
-15/+4	4	0	9	6	0	1	0	2
-16/+5	10	1	2	1	1	3	1	1
-17/+6	0	4	6	1	4	2		

23/34

FIG.20

24/34

FIG.21

A

B

25/34

FIG.22

Motif I (6a)

```
                UUGAGAAA    G
                             CAC U
5'...gGGUGCA    |||  |||     GUG U
3'...ucuaUGU                     U
        --CUCA-G
```

Motif II (1c)

```
              AAGAUA    UCU
                         CAGC    A
5'...AGAUG    |         GUCG    UAG
3'...ucuAC
        -ACA-G
```

Motif III (9a)

```
CC   UUGaucua-      ugaa-3'
A  GG  ||||        GCUU-5'
A  CC  ||||
UU   --CUAGUAA
```

WT (Motif II-like Domain)

```
         ACGGUA
     U            CA.....3'
     A  GACGCUG  ||
        CUGCGAC  ||  GU.....5'
     A       -GCG-G
```

FIG.23

FIG.24

28/34

FIG.25

% RNA Bound

[REV Protein] log₁₀(1 X 10⁻⁷ M)

C

17

12

13

6a

FIG.26

30/34

FIG.27

31/34

FIG.28

38/34

FIG.29

33/34

FIG.30

Motif II

31/34.

FIG. 31

INTERNATIONAL SEARCH REPORT

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) *

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5):C12Q 1/68;C12P 19/34;G01N 33/48,33/566;C07H 15/12
U.S. CL.: 435/6,91; 436/94,501;536/26,27,28,29

II. FIELDS SEARCHED

Minimum Documentation Searched 7

Classification System	Classification Symbols
U.S. CL.:	435/6,91; 436/94,501; 536/26,27, 28, 29; 935/77,78

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched 8

Computer Search: Dialog one search

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
Y	Belfort et al., eds. "RNA: Catalysis, Splicing. Evolution" published 1989 by Elsevier (Amsterdam), see pages 83-87.	1-39
Y	Biochemistry. Vol. 26. No. 6. issued 1987. Romaniule et al., "RNA Binding Site of R17 Coat Protein". pages 1563-1568, see entire paper.	1-39
Y	Biochemistry. vol. 22. no. 11. issued 1983. Carey et al., "Sequence Specific Interaction of R17 Coat Protein with its Ribonucleic Acid Binding Site". pages 2601-2610. See page 2603. column 1.	1-39
Y	Nucleic Acids Research. Vol. 17. No. 2. issued 1989. Joyce et al., "A Novel Technique for the Rapid Producing of Mutant RNAs". pages 711-722, see entire paper.	1-39

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
08 October 1991	21 OCT 1991

International Searching Authority	Signature of Authorized Officer
ISA/US	Stephanie W. Zitomer Ph.D. eb

Form PCT/ISA/210 (second sheet) (Rev.11-87)

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V.☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1.☐ Claim numbers   . because they relate to subject matter [12] not required to be searched by this Authority, namely:

2.☐ Claim numbers   .. , because they relate to parts of the international application that do not comply with the prescribed require-
ments to such an extent that no meaningful international search can be carried out [13], specifically:

3.☐ Claim numbers_____ , because they are dependent claims not drafted in accordance with the second and third sentences of
PCT Rule 6.4(a).

VI.☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

. This International Searching Authority found multiple inventions in this international application as follows:

    I- Claims 1-23, 30-39 comprise a first product and process of using.
    II- Claims 24-29 comprise a second process.
    III- Claims 40-90 comprise a second product.

1.☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
of the international application.

2.☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only
those claims of the international application for which fees were paid, specifically claims:

3.☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to
the invention first mentioned in the claims; it is covered by claim numbers:

4.☒ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not
invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (supplemental sheet (2) (Rev. 11-87,

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	Proceedings National Academy of Science. USA, Vol. 63. issued 1969. Cohen et al.. "Interactions of Hormonal Steriods with Nucleic Acids, i. A specific Requirement for Guanine". pages 458-464, see abstract.	40-90

THIS PAGE BLANK (USPTO)